

CHARACTERISATION OF NOVEL PROTEIN  
PARTNERS OF THE CCCTC-BINDING FACTOR IN  
BREAST CANCER CELL LINES

OKEZIE O OFOR

THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

AT

ANGLIA RUSKIN UNIVERSITY  
CAMBRIDGE, CHELMSFORD AND PETERBOROUGH

2015

## Acknowledgements

---

It is difficult to believe that this process has finally come to an end. The end looked so far away at the beginning yet it has come to pass.

The completion of this PhD involved many stakeholders and my gratitude goes first to God Almighty for the courage to embark on a PhD in the depths of the worldwide financial crisis. To my dad (Professor Ogbonnaya Ofor) and late father-in-law (Professor Obinna Okore Elendu) both of who never tired in their encouragement. I especially regret that my father-in-law is not here to see the final product of his prayers. To my parish priest, Rev Dr Karowei Dorgu, who let out a word of wisdom, 'take it day by day', in the bleakest times during the journey.

I owe a world of gratitude to my supervisors, Dr Christina Greenwood, Prof David Humber and Dr Helen Moor, all of who worked tirelessly on my incessant drafts. I still owe them a box of chocolates! To Dr France Docquier with whom I commenced the work originally and for all the arguments in the lab and to Professor Stephen Bustin whose input was especially valuable.

To all my Nigerian medical colleagues who thought I had truly gone mad to opt for a PhD while holding multiple medical qualifications.

Finally, I thank my wife Dr Nkechinyere Ofor, who very patiently held the fort in her capable hands and our children Chijioke and Onyeka both of whom have eternally wondered why Daddy could not take them to school over the past few years.

## ABSTRACT

CTCF is an evolutionally conserved 11-zinc finger protein. It controls multiple cellular functions and is itself partly regulated via poly (ADP-ribosyl)ation (PARylation). Recent evidence suggested that the hypo PARylated isoform was exclusively expressed in breast tumours and low expression levels was associated with worse prognostic features. It also possessed proliferative activity. The exact mechanism by which CTCF exerted its effects in breast cancer is not known.

In order to define the interaction between CTCF and proliferation markers, Ki67 and PCNA, colocalisation studies were performed in a panel of five breast cancer cell lines which possessed different hormone receptor / invasive phenotypes. Following co-localisation, co-immunoprecipitation assays and mass spectrometry were carried out to determine whether CTCF was physically bound to either Ki67 or PCNA. To determine whether CTCF directly regulated *ERα* activity, CTCF plasmid overexpression and siRNA knockdown assays were performed in the hormone receptor positive MCF7 breast cancer cell line. Changes in endogenous expression of *ERα* were monitored by quantitative polymerase chain reaction (QPCR).

CTCF, Ki67 and PCNA were all found to be strongly expressed in breast cancer cell lines though the strength of this expression for CTCF and Ki67 was antibody dependent. CTCF and Ki67 were shown to colocalise in the nucleoli of all breast cancer cell lines while CTCF and PCNA demonstrated nucleolar colocalisation only in the weakly invasive, hormone receptor positive cell lines. Despite colocalisation, there was no physical interaction detected between CTCF and Ki67 / PCNA on coimmunoprecipitation and mass spectrometry in all the cell lines studied suggesting that the proteins did not exist as a functional complex. Three novel CTCF - interacting protein partners (general transcriptional factor 2, glucose regulated protein 78 and the huntingtin interacting protein-1 related) were however discovered. These new protein partners, known to function at least in part via epidermal growth factor receptor (EGFR) signalling in cancer formation, were discovered only in ER positive breast cancer cell lines. Further investigation did not detect a direct regulatory effect of CTCF on *ERα* expression suggesting that the effect of CTCF on EGFR signalling in breast cancer cell lines did not involve an indirect action on *ERα* expression.

Key words: CTCF, breast cancer, Ki67, PCNA, estrogen receptor, protein partners.

## Table of Contents

Acknowledgements.....	i
Abstract.....	ii
Table of contents.....	iii
List of figures .....	xi
List of tables .....	xv
Copyright Declaration .....	xvi
Abbreviations .....	xvii
Chapter 1: Introduction.....	6
1.1 Functional anatomy of the female breast .....	6
1.2 Classification of breast cancer .....	9
1.3 Prognostic factors in breast cancer.....	16
1.4 Genetic alterations in breast cancer .....	19
1.5 Epigenetic alterations in breast cancer .....	21
1.6 The CCCTC-binding factor (CTCF) .....	23
1.6.1 History, discovery and conservation of CTCF.....	23
1.6.2 CTCF structure .....	24
1.6.3 CTCF functions.....	28
1.6.3.1. Genomic organiser and long-range chromatin interaction role .....	28
1.6.3.2 Regulation of gene expression .....	31
1.6.3.3 CTCF insulation and genomic imprinting role .....	31



1.6.3.4 X chromosome inactivation .....	34
1.6.3.5 Summary .....	34
1.7 CTCF DNA – binding sites.....	35
1.7.1 Distribution and characteristics of CTCF DNA - binding sites .....	35
1.7.2 Classification and functional effect of CTCF binding sites.....	37
1.8 CTCF protein partners.....	38
1.8.1 Chromatin CTCF - protein partners.....	40
1.8.2 DNA-binding CTCF protein partners .....	40
1.8.3 Multifunctional CTCF protein partners .....	41
1.8.4 Miscellaneous CTCF protein partners.....	42
1.8.5 Summary .....	42
1.9 Regulation of CTCF activity.....	42
1.9.1 Methylation.....	43
1.9.2 Post translational modifications (PTM).....	43
1.9.2.1 PARylation .....	43
1.9.2.2 SUMOylation .....	44
1.9.2.3 Phosphorylation .....	45
1.10 CTCF and tumour suppression.....	45
1.10.1 CTCF induces transcriptional repression of hTERT in tumours .....	45
1.10.2 CTCF maintains retinoblastoma protein (pRb) and p53 gene promoter epigenetic status and tumour repression.....	46

1.10.3 CTCF maintains epigenetic balance at the cyclin-dependent kinase inhibitor 2A locus (CDKN2A).....	47
1.11 CTCF and the clinical breast cancer phenotype .....	48
1.12 CTCF and the estrogen receptor $\alpha$ .....	51
1.13 Conclusion.....	53
1.14 Aims and Objectives.....	54
Chapter 2: Materials and Methods .....	55
2.1 MATERIALS .....	55
2.1.1 Breast cancer cell (BCC) cell lines .....	55
2.1.2 Culture media.....	55
2.1.2.1 Culture medium for MCF7, T47D and BT474 cells.....	55
2.1.2.2 Culture medium for SKBR3 and MDA MB 231 cells .....	55
2.1.2.3 culture medium for LDM226 breast cells .....	56
2.1.2.4 Luria Bertani broth (LB) and Luria agar .....	56
2.1.3 Reagents and Buffers /gels / solutions.....	56
2.1.4 Antibodies .....	56
2.1.5 Plasmids, siRNA and biologic agents .....	56
2.1.6 QPCR primers.....	56
2.2 METHODS .....	58
2.2.1. Cell culture procedures.....	58
2.2.1.1 Reviving cells from storage .....	58
2.2.1.2 Cell passaging and cell count .....	58

2.2.1.3 Freezing down cells .....	58
2.2.2 Trypan blue test for cell viability .....	59
2.2.3 Breast cancer cell lysates.....	59
2.2.3.1 Cell lysate for Western blot analysis .....	59
2.2.3.2 Cell extract for immunoprecipitation .....	59
2.2.4 Bovine serum albumin protein assay.....	60
2.2.5 Indirect Immunofluorescence procedure.....	60
2.2.6 Sodium Dodecyl Sulphate (SDS) – PolyAcrylamide Gel Electrophoresis (PAGE) and western blot analysis .....	62
2.2.6.1 Gel preparation and electrophoresis .....	62
2.2.6.2 Semi-dry transfer .....	62
2.2.6.3 Blocking, primary and secondary antibody incubations .....	63
2.2.6.4 Blot development .....	63
2.2.6.5 Stripping blot membranes.....	64
2.2.6.6 Silver staining .....	64
2.2.6.7 Coomassie blue staining .....	65
2.2.7 Immunoprecipitation assay (IP).....	65
2.2.8 RNA – based procedures .....	66
2.2.8.1 RNA extraction from cells .....	66
2.2.8.2 RNA integrity assessment using the Agilent 6000 Bioanalyser .....	67
2.2.9 Plasmid DNA procedures .....	68

2.2.9.1 Bacterial cell transformation using DDH5α <sup>™</sup> competent cells .....	68
2.2.9.2. Preparation of bacterial culture .....	68
2.2.9.3 Plasmid DNA minipurification .....	68
2.2.9.4 DNA quantitation with UV spectrophotometry (Nanodrop) .....	70
2.9.5 Plasmid restriction enzyme digestion and agarose gel electrophoresis.....	70
2.2.9.6 Plasmid extraction using the Endofree plasmid Maxiprep kit (Qiagen <sup>™</sup> ).....	70
2.2.10 Transfection assays.....	71
2.2.10.1 Transfection assays with plasmid expression vectors .....	71
2.2.10.2 Transfection assays with small interfering RNA (siRNA) .....	72
2.2.11 Reverse transcription – polymerase chain reaction (RT-PCR; QPCR) procedures .....	73
2.2.11.1 Complimentary DNA (cDNA) synthesis .....	73
2.2.11.2 Preparing primers for QPCR .....	73
2.2.11.3 Standard curve determination for QPCR efficiency .....	73
2.2.11.4 Quantitative polymerase chain reaction (QPCR) .....	73
2.2.12 Liquid chromatography – mass spectrometry (LC – MS / MS).....	74
Chapter 3. RESULTS - Investigating CTCF protein partners in a panel of five breast cancer cell lines . .....	75
3.1 Background .....	75
3.1.1 Cell proliferation and breast cancer .....	75

3.1.2 CTCF and proliferation .....	75
3.1.3 Ki67 protein and proliferation .....	77
3.1.4 Proliferating cell nuclear antigen (PCNA) protein, proliferation and cancer.....	79
3.2 Knowledge gap and hypothesis .....	80
3.3 Objectives of this chapter .....	80
3.4 Results .....	81
3.4.1 Confirmation of estrogen receptor (ER), progesterone receptor (PR) and HER2 receptor expression status in breast cancer cell lines .....	81
3.4.2 Total CTCF protein expression in different breast cancer cell lines .....	83
3.4.3 Differential expression of proliferation markers, Ki67 and PCNA, in a panel of different breast cancer cell lines .....	86
3.4.4 CTCF localisation in relation to breast cancer phenotype and anti-CTCF antibody type .....	89
3.4.5 Ki67 protein localisation in a panel of five breast cancer cell lines using two different anti Ki67 antibodies .....	93
3.4.6 PCNA protein localisation in breast cancer cell lines .....	97
3.4.7 Assessment of immunofluorescence bleed through .....	100
3.4.8 CTCF protein co-localisation with Ki67 protein in breast cancer cell lines.....	102
3.4.9 CTCF protein colocalisation with PCNA protein in breast cancer cell lines with different hormone receptor / HER2 and invasive properties.....	106

3.4.10 CTCF immunoprecipitation (IP) and coimmunoprecipitation (co-IP) with Ki67 and PCNA in MCF7 breast cancer cells using a high stringency IP buffer.....	111
3.4.11 CTCF co-immunoprecipitation with Ki67, PCNA and known protein partners (RNA pol II and PARP 1) in MCF7 breast cancer cells using a medium stringency buffer. ....	114
3.4.12. CTCF immunoprecipitation and coprecipitation with Ki67, PCNA and known protein partners (RNA pol II and PARP 1) in MCF7 breast cancer cells using a low stringency buffer.....	116
3.4.13. PCNA immunoprecipitation and coprecipitation with CTCF and Ki67 in the MCF7 breast cancer cell line using a low stringency IP lysis buffer.....	119
3.4.14 Immunoprecipitation of CTCF and coprecipitation with Ki67 / PCNA using a cross-linker and a magnetic commercial kit .....	121
3.4.15 RNA pol II immunoprecipitation and coprecipitation with CTCF in the HeLa cervical cancer cell line cell using a low stringency IP lysis buffer (buffer 3).....	123
3.4.16 Breast cancer cell line - specific differences in CTCF coprecipitation with Ki67 and PCNA in breast cancer cell lines .....	125
3.4.17 CTCF protein interaction with Ki67 / PCNA assessed by Liquid Chromatography - Mass Spectrometry (LC – MS / MS) in breast cancer cell lines.. ....	128
3.4.18 Novel CTCF-interacting protein partners in breast cancer cell lines .....	132
3.4.19 CTCF interacting protein network map in breast cancer cell lines .....	135
Chapter 4. Discussion: Novel CTCF interacting partners.....	143
Chapter 5. Regulatory relationship between CTCF and ER $\alpha$ .....	157

5 Background .....	157
5.1 Role of estrogen in breast cancer initiation.....	157
5.2 ER structure and activation .....	160
5.3 ER $\alpha$ regulation .....	162
5.4 Knowledge gap and hypothesis .....	164
5.5 Objectives of this chapter .....	164
5.6 Results .....	165
5.6.1 Restriction enzyme digest of CTCF plasmid expression vectors .....	165
5.6.2 Optimisation of CTCF plasmid DNA overexpression assays in MCF7 cells .....	169
5.6.3 Assessment of siRNA transfection efficiency and experimental set up with the positive control cyclophilin siRNA .....	172
5.6.4 Optimisation of CTCF siRNA knockdown assays in MCF7 cells.....	174
5.6.5 Effect of transfection agents and reagents on MCF7 cell growth and viability. ....	177
5.6.6 QPCR measurements of CTCF and ER $\alpha$ gene expression following plasmid overexpression and siRNA knockdown assays in MCF7 cells .....	181
5.6.6.1 Bioanalyser and Nanodrop estimation of RNA quality and concentrations .....	182
5.6.6.2 DINAmelt <sup>R</sup> prediction of secondary structure of primer pairs for QPCR.....	187
5.6.6.3 QPCR assay optimisation: linear range and dilution factor cDNA .....	190
5.6.6.4 QPCR optimisation: standard curves and melting curves .....	193

5.6.6.5 Variation in CTCF mRNA expression and ER expression response in MCF7 breast cancer cells .....	198
Chapter 6. Discussion – Regulatory effect of CTCF on ER $\alpha$ expression.....	203
Chapter 7. Final discussion and conclusions .....	208
FUTURE DIRECTIONS.....	212
REFERENCES.....	213
INTERNET SITES.....	247
APPENDIX .....	248



## List of Figures

Figure 1.1 European age-standardised incidence rates per 100,000 population, by age, females, Great Britain .....	2
Figure 1.2. Breast Cancer mortality: 2010 – 2012, United Kingdom.....	5
Figure 1.3 Anatomy of the normal female breast .....	7
Figure 1.4 Schematic representation of a transverse section through a normal breast duct. ....	8
Figure 1.5. Histology of breast ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS).....	11
Figure 1.6. Histology of invasive ductal (IDC) and invasive lobular carcinoma (ILC) .....	12
Figure 1.7. Human breast tumour subtypes and possible link to normal breast epithelial hierarchy .....	13
Figure 1.8 Structural features of CTCF.....	26
Figure 1.9. Schematic structure of the classic Cys2 His2 zinc finger (ZF) .....	27
Figure 1.10. CTCF and imprinting at the Igf2 / H19 gene locus via loop formation .....	33
Figure 1.11. Schematic representation of classes of CTCF protein partners.....	39
Figure 3.1: Hormone receptor (ER and PR) and HER2 expression profile of five breast cancer cell lines .....	82
Figure 3.2: Total CTCF protein expression in a panel of five breast cancer cell lines and one normal breast epithelial cell line .....	85
Figure 3.3: Expression of proliferation markers Ki67 and PCNA in a panel of five different breast cancer cell lines and one normal breast epithelial cell line .....	88
Figure 3.4. Single indirect immunofluorescence staining of five breast cancer cell lines with CTCF primary antibodies .....	91
Figure 3.5. Single immunofluorescence staining of five breast cancer cell lines with Ki67 primary antibodies .....	94 - 95

Figure 3.6. Single immunofluorescence staining of five breast cancer cell lines with PCNA primary antibody .....	98
Figure 3.7 Assessment of fluorophore-conjugated secondary antibody bleedthrough.....	101
Figure 3.8 Double immunofluorescence of CTCF protein with Ki67 protein in a panel of five breast cancer cell lines .....	103
Figure 3.9 A - E. Double indirect immunofluorescence of CTCF with PCNA in a panel of five breast cancer cell lines .....	107
Figure 3.10. Silver stain (A), CTCF immunoprecipitation (B) and CTCF co-immunoprecipitation (C) in MCF7 cells using a high stringency buffer .....	113
Figure 3.11. CTCF co-IP with Ki67, PCNA, RNA pol II, and PARP 1 in MCF7 breast cancer cells using a medium stringency IP buffer .....	115
Figure 3.12. CTCF IP and co-IP with Ki67 / PCNA and known protein partners RNA pol II / PARP 1, in MCF7 breast cancer cells using a low stringency buffer .....	118
Figure 3.13. PCNA immunoprecipitation and coprecipitation with CTCF / Ki67 in MCF7 breast cancer cells using a low stringency IP lysis buffer .....	120
Figure 3.14. CTCF immunoprecipitation (IP) and co-immunoprecipitation (co-IP) in MCF7 cells using a magnetic co-IP kit incorporating a crosslinker .....	122
Figure 3.15. RNA pol II and CTCF coimmunoprecipitation in HeLa cell extracts .....	124
Figure 3.16. A - D. CTCF IP and co-IP with Ki67 / PCNA in T47D, BT474, SKBR3 and MDA MB 231 breast cancer cell lines .....	126
Figure 3.17. CTCF spectral scores on liquid chromatography / mass spectrometry in breast cancer cell lines .....	130
Figure 3.18. CTCF-interacting proteins identified across all cell lines .....	134
Figure 3.19. Action view of CTCF interacting proteins identified in at least two breast cancer cell lines with SC of 2 or more .....	137
Figure 3.20. STRING output of known and predicted CTCF interacting protein partners in homo sapiens .....	138

Figure 3.21 A – D. Action view of breast cancer cell line - specific CTCF protein interaction network .....	139
Figure 5.1. Inter-relation of estrogen, ER and EGFR receptor signalling cascades.....	158
Figure 5.2. Structure and comparison of estrogen $\alpha$ and $\beta$ receptor domains .....	161
Figure 5.3. Restriction enzyme digest of CTCF pCI expression vector .....	167
Figure 5.4. Restriction enzyme digest of the pCI empty vector (EV).....	168
Figure 5.5. Optimisation of MCF7 overexpression with a plasmid expression vector.....	171
Figure 5.6. CTCF siRNA transfection efficiency in MCF7 cells with cyclophilin B positive control .....	173
Figure 5.7. MCF7 cell RNA interference (RNAi) with CTCF and non target siRNA .....	176
Figure 5.8. MCF7 cell response to expression vectors and attractene transfection reagent .....	179
Figure 5.9. MCF7 cell response to siRNA and DharmaFECT transfection reagent .....	180
Figure 5.10. Densitometry plot of RNA quality assessment using the Agilent Bioanalyser .....	184
Figure 5.11. Spherograms of RNA quality assessment using the Agilent Bioanalyser.....	185
Figure 5.12. Nucleic acid primer pair fold pattern for CTCF and ER $\alpha$ primer pairs.....	188
Figure 5.13. Nucleic acid primer pair fold pattern for TBP and GAPDH primer pairs.....	189
Figure 5.14. Testing primers: amplification plots for linear range.....	191
Figure 5.15. Standard curves for QPR efficiency .....	195
Figure 5.16. Assessment of QPCR replicate consistency and melting curves .....	197
Figure 5.17. CTCF and ER $\alpha$ mRNA fold change on CTCF overexpression.....	200
Figure 5.18. CTCF and ER mRNA fold change on siRNA knockdown .....	202

## List of Tables

Table 1.1 Anatomical classification and prevalence of breast cancer subtypes (worldwide).....	14
Table 1.2. Classification of breast cancer based on gene expression profile.....	15
Table 1.3. Prognostic factors in breast cancer .....	18
Table 1.4. Categories of CTCF chromatin loop domains.....	30
Table 2.1 Tabular annotation of breast cancer cell lines used in this thesis .....	57
Table 3.1. Relative expression levels of Ki67 protein in five breast cancer cell lines.....	96
Table 3.2. Tabular annotation of expression patterns of CTCF, Ki67 and PCNA proteins by single immunofluorescence in breast cancer cell lines .....	99
Table 3.3. Summary of colocalisation of CTCF with Ki67 and PCNA in a panel of five breast cancer cell lines.....	110
Table 3.4. Distribution of total number of CTCF-interacting proteins identified on label-free mass spectrometry .....	131
Table 3.5 Tabular annotation of novel CTCF interacting partners detected by mass spectrometry .....	133
Table 5.1. Bioanalyser and Nanodrop spectrometric estimation of RNA concentrations and purity .....	186
Table 5.2. Tabular annotation of Cq values of amplification plots for linear range .....	192
Table 5.3: QPCR of plasmid vector overexpression assays, Cqs and fold change .....	199
Table 5.4: Cq values and fold change for CTCF siRNA knockdown assays ... ..	201

**Copyright Declaration**

I, Okezie O Ofor, confirm that the work presented in this thesis is my own. Where information has been taken from other sources, I confirm that this has been referenced appropriately.

## Abbreviations

HER2	Human epidermal receptor 2
WHO	World Health Organisation
CCCTC-binding factor	CTCF
DCIS	Ductal carcinoma in-situ
LCIS	Lobular carcinoma in-situ
ER	Estrogen receptor
PR	Progesterone receptor
TNM	Tumour, Node, Metastasis
EGFR	Epidermal growth factor receptor
TNBC	Triple negative breast cancer
<i>BRCA 1</i>	Breast Cancer 1
<i>BRCA 2</i>	Breast Cancer 2
AR	Androgen receptor
TSG	Tumour suppressor gene
TSS	Transcription start site
ZF	Zinc finger
CHIA - PET	Chromatin immunoprecipitation and paired end tag extraction and sequencing
ES	Embryonic stem (ES)
ChIP-Seq	Chromatin immunoprecipitation – sequencing
TF	Transcriptional factor
MHCII	Major histocompatibility complex class II
3C-Seq	Chromosome conformational capture - sequencing
ICR	Imprint control region
DMR1	DNA methylated region 1
bp	base pair
PWM	Positional weight matrix
EGFR	Epidermal growth factor receptor
PARP 1	Poly ADP ribose polymerase 1
RNA pol II	RNA polymerase II

PTM	post translational modification
PARylation	Poly (ADP-ribosyl) ation
PAR	Poly-ADP ribose
PARPs	Poly-ADP ribose polymerases
SUMO	Small Ubiquitin-like Modifier
hTERT	Human telomerase reverse transcriptase
HDAC	Histone deacetylase
pRb	Retinoblastoma protein
AZA	5'-aza-2'-deoxycytidine
RT-PCR	Real time polymerase chain reaction
E2	Exogenous estrogen
ERE	Estrogen response element
QPCR for	Reverse transcription – quantitative polymerase chain reaction
EDTA	Ethylene diamine tetra acetic acid
DMSO	Dimethylsulphoxide
IP	Immunoprecipitation
PVDF	Polyvinylidene fluoride
HRP	Horseradish peroxidase
ECL	Enhanced chemiluminescence
Rpm	Revolutions per minute
siRNA	Small interfering RNA
cDNA	Complimentary DNA
LC – MS / MS	Liquid chromatography – mass spectrometry
rRNA	Ribosomal RNA
IBCSG	International Breast Cancer Study Group
POETIC	Trial of Perioperative Endocrine Therapy – Individualising Care
OS	Overall survival
DFS	Disease free survival
ATTC	American Tissue Type Collection
SDS-PAGE	Sodium dodecyl poly acrylamide gel electrophoresis

IHC	Immunohistochemistry
WB	Western blotting
STRING	Search Tool for the Retrieval of INteracting Genes / Proteins
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR experiments
smFISH	single molecule fluorescence <i>in situ</i> hybridization

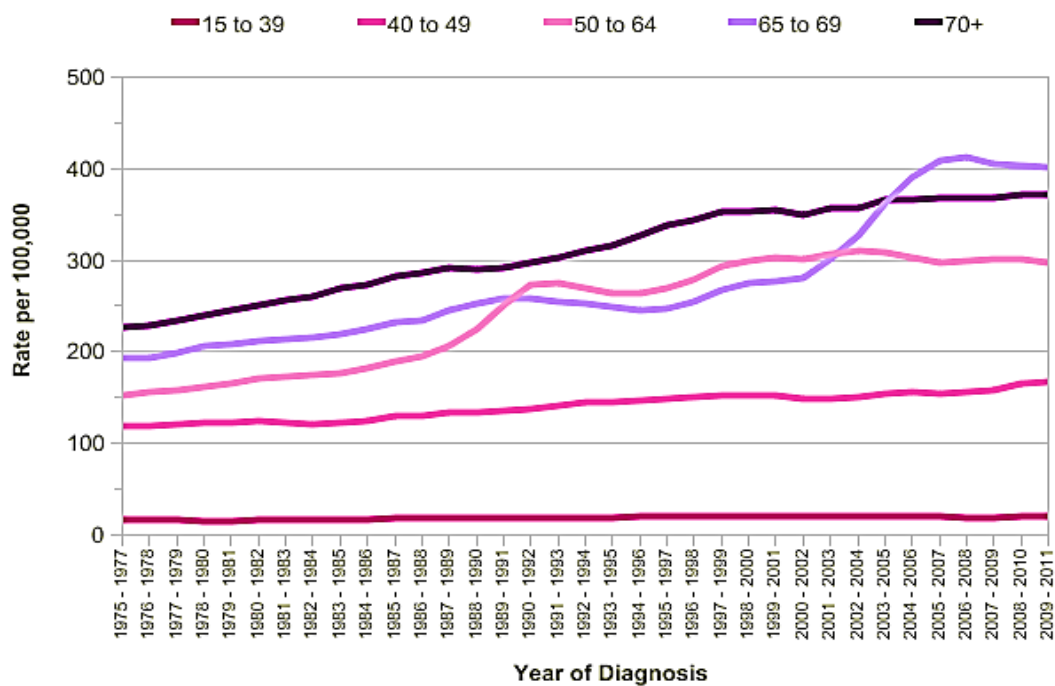


# Chapter 1

---

## Introduction

Breast cancer is a disease characterised by the malignant transformation of normal breast cells and in 2011 comprised 30% of new cancers in females in the United Kingdom (Office for National Statistics, Cancer Statistics Registrations, England [Series MB1]). Women in the United Kingdom have a 1 in 8 lifetime risk of developing the disease (Cancer Research UK, Stats Terminology and Calculations). Risk factors for breast cancer include early menarche, late menopause, reduced parity and lack of breast feeding which suggest that sustained and unopposed hyperestrogenism is important in breast cell proliferation and transformation (Jemal *et al.*, 2011). Germ line mutations and family history of breast cancer also contribute to the risk of breast cancer and suggest the influence of primary gene disorders (Barrett, 2010). When people move from regions of deprivation to those of greater affluence and from countries of low risk to those of higher risk, then there is a higher chance of breast cancer occurring suggesting a significant environmental component (Maringe *et al.*, 2013). The risk for breast cancer also increases with age, as 81% of women in England who develop the disease are diagnosed after the age of 50 years and about half of these diagnosed cases occur in women aged between 50 and 69 years (Barrett, 2010). The age-specific incidence of breast cancer is rising steadily as shown in figure 1.1 and may be attributable to population-wide screening and frequent self-breast examination leading to earlier diagnosis (Pocobelli and Weiss, 2014). Though breast cancer incidence is rising, survival rates have improved over the past four decades with 5-year survival rate improving from 50% in the 1970s to 85.8% (between 2008 and 2012) in English patients (Office for National Statistics, Cancer Statistics Registrations, England). This improvement in survival has been attributed not only to the earlier mentioned progress

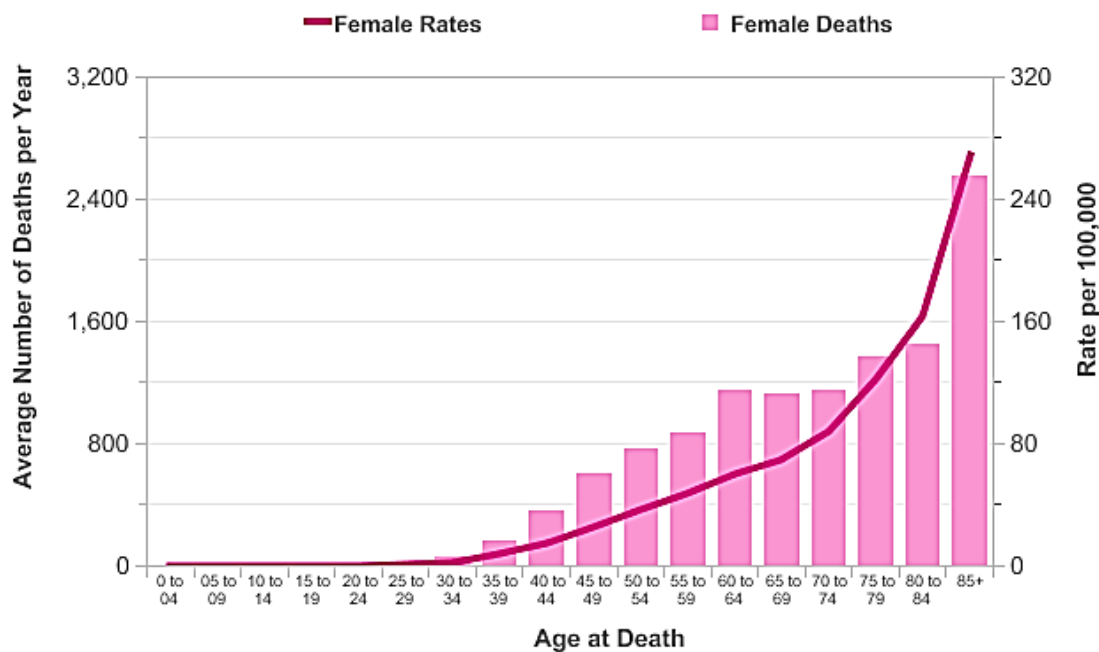


**Figure 1.1. European age-standardised incidence rates per 100,000 population, by age, females, Great Britain.** The different colour-coded lines trace the incidence of breast cancer across all age groups from 1975 to 2011. Aside from the 15 to 39 age group there is a persistent rise in breast cancer incidence across all age groups in the period under study. Source: Cancer Research UK, Stats Terminology and Calculations.

achieved in early diagnosis but also the evolution of multiple treatment options which include surgery, radiotherapy, chemotherapy, endocrine therapy and the application of new targeted biologic therapies (Bilal *et al.*, 2013). With respect to treatment, local control of breast cancer involves surgery and radiotherapy. For those breast tumours treated by surgical removal, the target is to have clear surgical resection margins defined as 'no ink on the surgical specimen' which refers to the absence of cancerous cells at the margins of the excised tumour (Moran *et al.*, 2014). Radiotherapy also offers tumour bed control further diminishing the risk of tumour recurrence (Moran *et al.*, 2014). Systemic treatments involve the administration of medications and serve to destroy or inhibit cancer cells that have migrated from the local breast area of origin. They include chemotherapy, endocrine agents and targeted biologic therapies. There are multiple lines of chemotherapy treatment which could be used singly or in combination and include 5-fluorouracil, epirubicin, cyclophosphamide, docetaxel, paclitaxel and capecitabine (reviewed in Miller *et al.*, 2014). Other systemic treatment options include endocrine drugs for those cancers that express the estrogen receptor (ER) and progesterone receptor (PR). These agents serve to inhibit hormone receptors and prevent the growth of the cancer and include agents like the anti-estrogen tamoxifen and aromatase inhibitors like anastrozole, letrozole, exemestane and fulvestrant (reviewed in Miller *et al.*, 2014). As with chemotherapy, the anti-hormonal agents are used sequentially when there is evidence that the cancer has become resistant to a particular anti-hormonal drug. The more recent advent of targeted therapies like tyrosine kinase inhibitors and monoclonal antibodies has improved prognosis, prediction and survival for breast cancer patients. The tyrosine kinase inhibitor eribulin is now used as a third line systemic treatment in the metastatic breast cancer setting (Twelves *et al.*, 2014). For breast cancers that overexpress the human epidermal receptor 2 (HER2), the monoclonal antibody, trastuzumab, has changed the prognosis of that group of breast cancers. In the adjuvant setting, it is recommended for use for a year after surgery and further decreases the risk of recurrence (Piccart-Gebhart *et al.*, 2005). In the metastatic

setting, there is more recent evidence suggesting that the addition of another monoclonal antibody, pertuzumab, to the trastuzumab plus chemotherapy combination in patients with HER2 positive metastatic breast cancer, provided an additional 15.7 months of overall survival on top of the 40 months achievable with trastuzumab plus chemotherapy alone (Swain *et al.*, 2014).

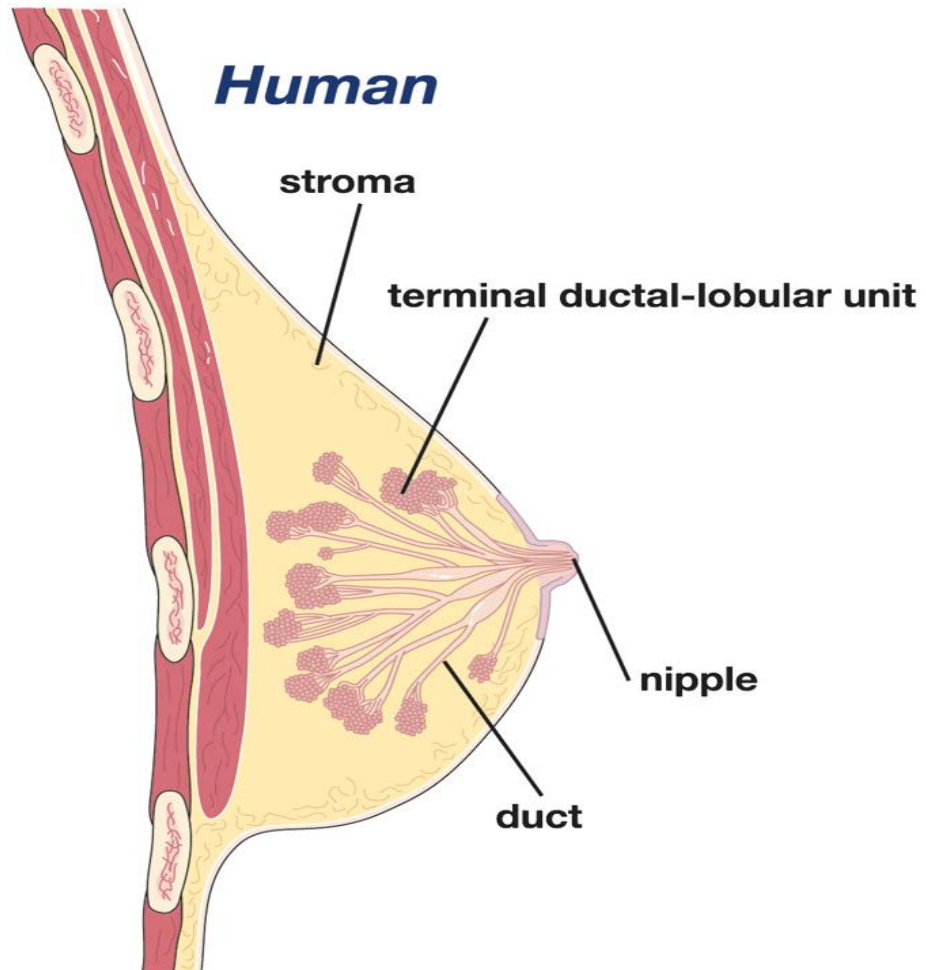
Despite these improvements breast cancer caused 519,000 deaths in 2004 and 460,000 deaths in 2008 on a global scale confirming that mortality from this condition remained high (WHO Fact sheet number 297). In 2012, breast cancer was second only to lung cancer as a cause of cancer death in the UK with 11, 716 deaths and had a mortality rate of 36 for every 100, 000 female (Cancer Research UK, Stats Terminology and Calculations). The age-specific mortality rate for breast cancer in the UK shown in figure 1.2 revealed a sustained rise in mortality with advancing age and breast cancer was also found to represent the commonest cause of cancer death in women aged 15 – 49 years (Cancer Research UK, Stats Terminology and Calculations). Contributing to breast cancer morbidity and mortality is the triple negative subset that does not express hormone receptors or overexpress HER2 and represent 10% – 15% of breast cancers. They can only be treated with chemotherapy as they possess no clear biologic targets and demonstrate a higher recurrence and poorer survival rates when compared with other types of breast cancer at a similar stage of disease (reviewed in Miller *et al.*, 2014). On account of the immense burden of breast cancer and the lack of targeted therapeutics for some subsets, there is therefore the urgent and continuing need for new and innovative efforts in translational research to discover and further widen the pool of available agents and knowledge in the fight against the disease. In an effort to contribute to that pool of knowledge, the information laid out in this thesis described the investigation of the mechanistic involvement of a multivalent protein, the CCCTC-binding factor (CTCF), in breast cancer as it had previously been discovered to be relevant in the breast cancer phenotype (Butcher and Rodenhiser, 2007; Docquier *et al.*, 2009).



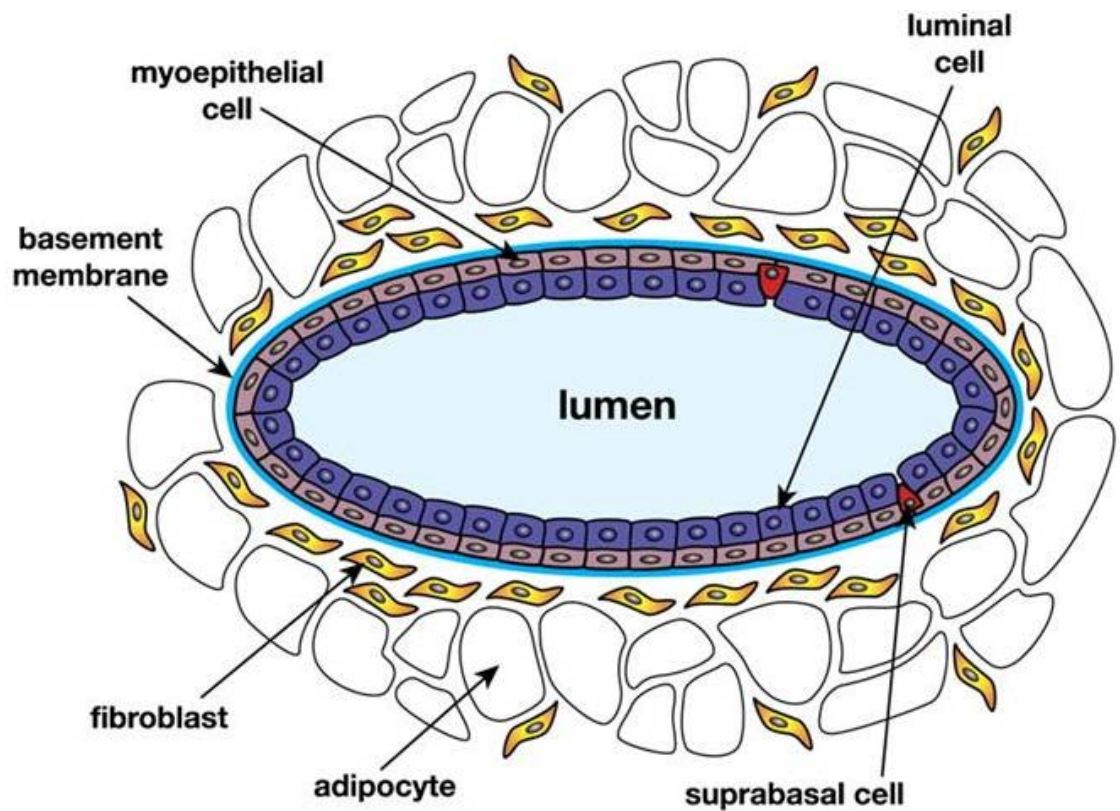
**Figure 1.2. Breast Cancer mortality: 2010 – 2012, United Kingdom.** Average number of deaths per year and age-specific mortality rates per 100,000 populations. The bars represent absolute death rates for each age group. The line in violet tracks the increase in death rate with advancing age. Source: Cancer Research UK, Stats Terminology and Calculations.

## 1.1 Functional anatomy of the female breast

The normal female breast is shown in figure 1.3 and comprises 15 to 20 functional lobes composed of fibrofatty tissue, gland lobules and ductal elements that coalesce to form the lactiferous ducts which terminate at the nipple (Jesinger, 2014). The lactiferous duct lumen is lined by columnar epithelium becoming squamous as the duct opens on the areola (Hoda, 2012). Shown in figure 1.4 are suprabasal cells scattered within the epithelial layer of the lactiferous ducts and between the epithelium and the basal lamina lie myoepithelial cells. Myoepithelial cells are specialized, contractile elements and are involved in the contractile responses of the breast during lactation. The lactiferous ducts are surrounded by smooth muscle which is in turn embedded in hormone - sensitive stroma and the ducts thin out as they extend from the nipple towards the chest wall (Hoda, 2012). The relative composition of the breast with respect to fibrofatty tissue, gland lobules and ductal elements is dependent on age and functional state (for instance pregnancy). With increasing age, glandular tissue is replaced by fibrofatty material. Functionally, on account of the preponderance of glandular tissue in the young female, investigating the breast with mammography in this age group leads to a high rate of false results (Engelken *et al.*, 2012). Any part of the normal breast - glandular lobules, ductal cells and stromal elements - could undergo malignant transformation. There are significant differences with respect to prevalence, prognosis and clinical behaviour between tumours arising from different parts of the breast (Melchor *et al.*, 2014). These differences characterise the heterogeneity of breast cancer and is rightly regarded as a group of diseases arising from the breast rather than a single clinical disease state. Breast cancer subdivision therefore reflects this heterogeneity.



**Figure 1.3 Anatomy of the normal female breast.** The various components and anatomical boundaries of the female breast in sagittal section are shown. Source: Visvader, 2009.



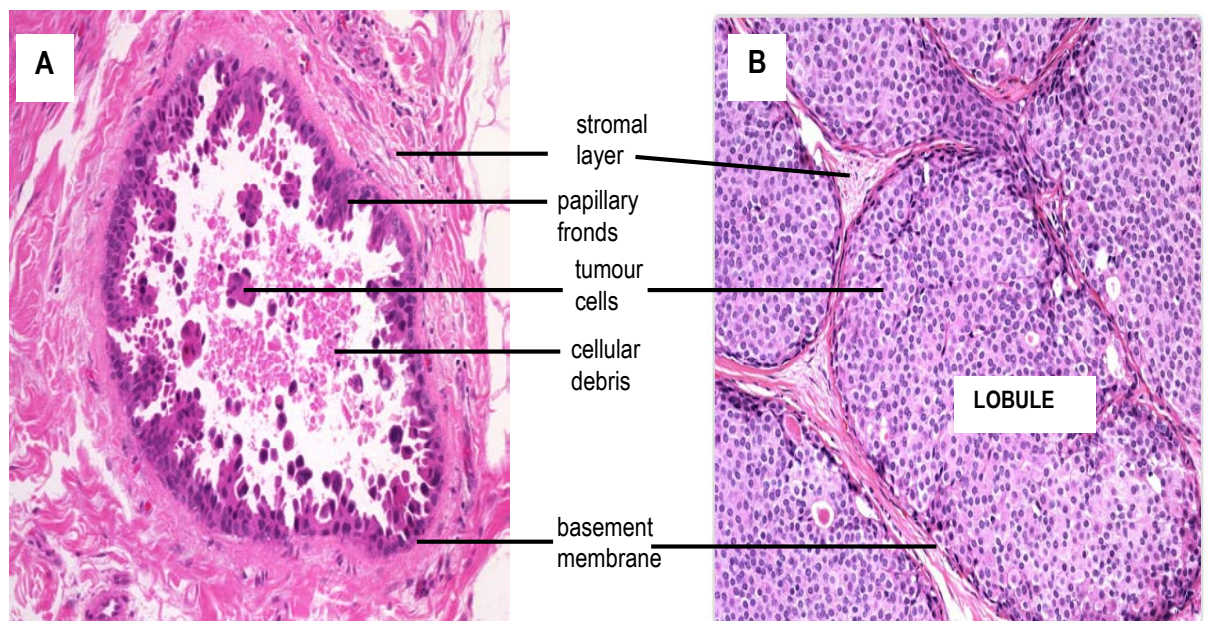
**Figure 1.4. Schematic representation of a transverse section through a normal breast duct.** A suprabasal cell sits on the myoepithelial layer but does not reach the lumen. Source: Visvader, 2009.



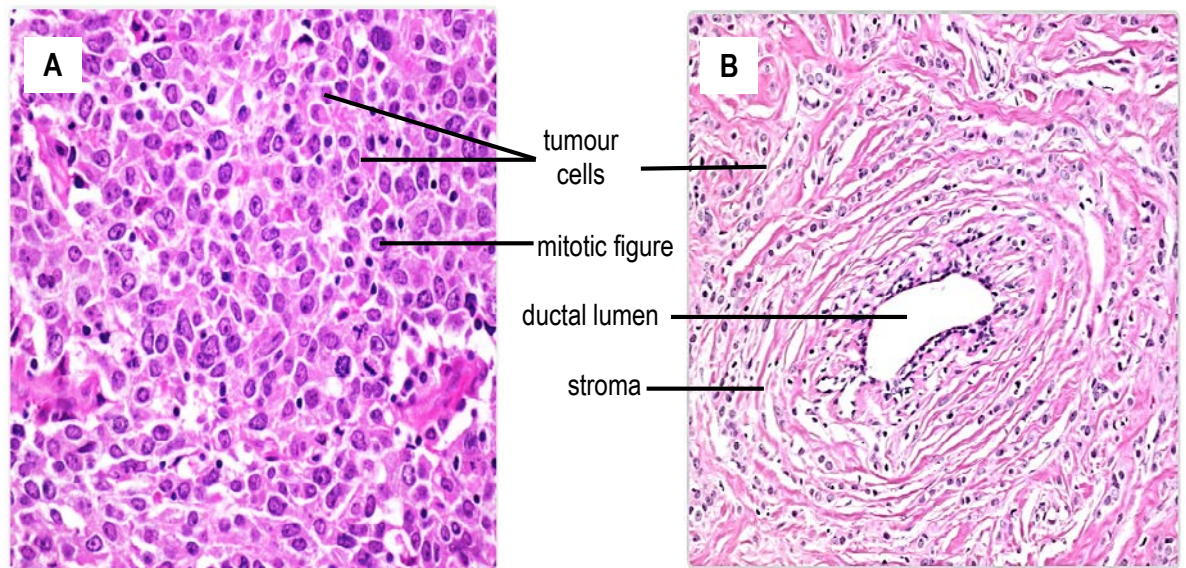
## 1.2 Classification of Breast Cancer

There are multiple classifications for clinical human breast cancer. Based on histology, breast cancer could be pre-invasive or invasive. Pre-invasive forms describe cellular transformation in the ductal or lobular cells with the basement membrane of transformed cells still intact and are termed ductal carcinoma in-situ (DCIS) or lobular carcinoma in-situ (LCIS) respectively (Geyer *et al.*, 2012; Rakha and Ellis, 2012). The histology of these forms of breast cancer are shown in figure 1.5 and reveal a form of DCIS with papillary extensions of tumour cells that project into the duct lumen while LCIS shows lobules infiltrated with round monochromatic tumour cells. In this Figure both ductal and lobular basement membranes remain intact. It is important to identify and treat preinvasive forms of breast cancer, as they are the harbinger of the invasive forms but possess better prognoses (Geyer *et al.*, 2012; Rakho and Ellis, 2012). The histologies of invasive ductal (A) and invasive lobular (B) breast carcinoma are shown in figure 1.6. This figure reveals pleomorphic tumour cells, extensive infiltration of the stroma, absence of an intact basement membrane and the generally disordered cellular architecture that distinguishes invasive from non-invasive breast cancers. Further classification of breast cancer can be based on the possible anatomic origin of the transformed cell in breast tissue. Figure 1.7 shows the presumed hierarchical association between breast cancer groups and their normal breast epithelial counterpart. Further summarised in table 1.1 are the different types of breast cancer, their anatomic origin in the breast and the prevalence of the different types in patients with breast cancer. The exact breast cell of origin of breast cancer however is not always clear as exemplified by the tubular, medullary cell and adenoid cystic forms (table 1.1). The medullary class of breast cancer is usually identified histologically where the tumour is clearly demarcated from normal breast tissue; while the tubular forms lack a myoepithelial layer but possess well-formed tubules directly in contact with the stroma (Debnath and Brugge, 2005). This anatomic classification nevertheless is important as

the tubular and lobular forms have better prognoses than the ductal type and need to be clearly identified (Debnath and Brugge, 2005). Further differentiation between classes of breast cancer based on gene expression analyses is provided in table 1.2. This table also shows the hormone receptor profile of each breast cancer group and their clinical characteristics. The hormonal profile refers to the expression of estrogen and progesterone receptors and the human epithelial receptor (HER) 2 by the tumour. As noted in table 1.2, the clinical characteristic of each tumour, a product of the aggregate hormone receptor profile, is extremely important, as it could be predictive for response to treatment options.

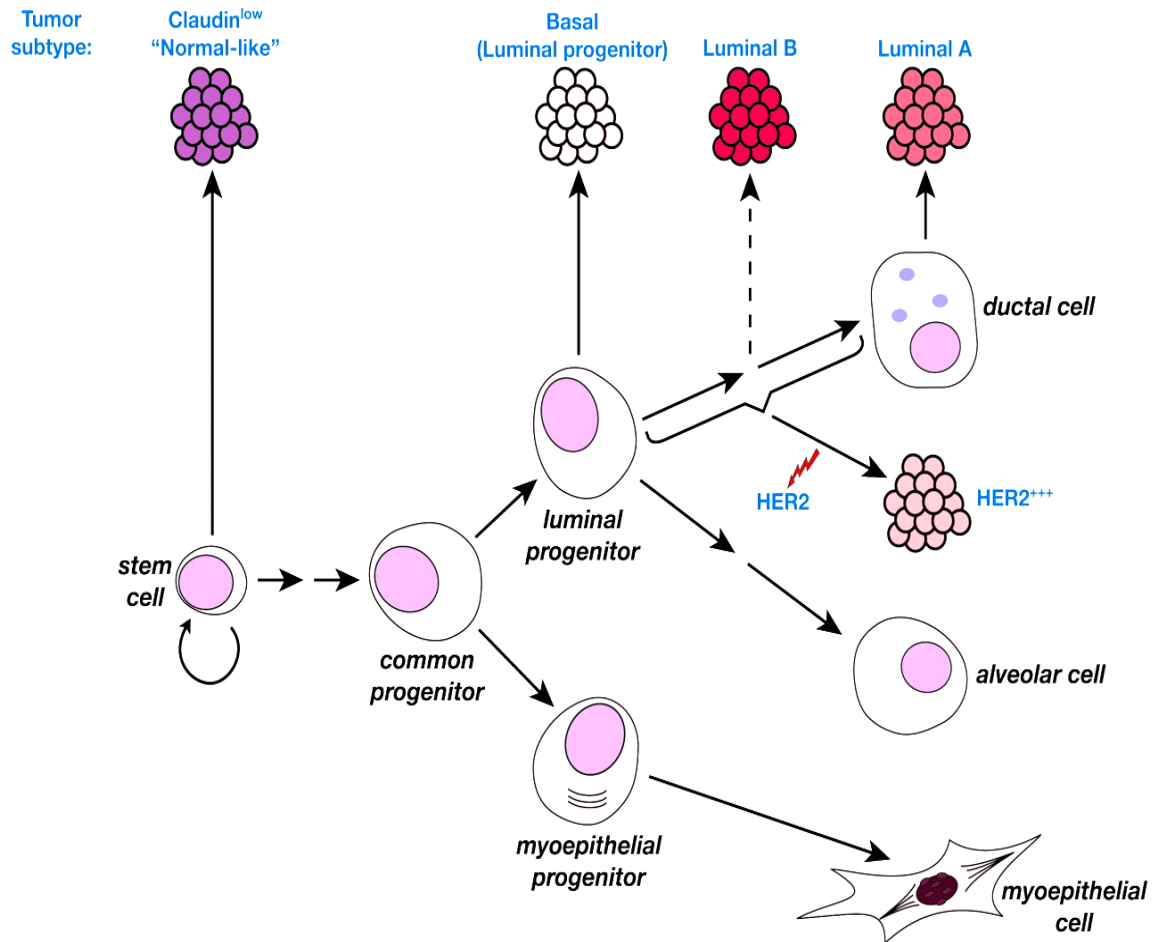


**Figure 1.5. Histology of breast ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS).** Hematoxylin and eosin stained cross-section of breast cancer showing (A) micropapillary form of DCIS with projections of papillary structures of different shapes bearing tumour cells. Cellular debris and detached clusters of tumour cells that show high nuclear grade are scattered in the ductal lumen. (B) LCIS with round uniform and monochromatic tumour cells in the breast lobule. The basement membranes in both A and B are intact. Source: WebPathology: Visual survey of surgical pathology.



**Figure 1.6. Histology of invasive ductal (IDC) and invasive lobular carcinoma (ILC).**

Hematoxylin and eosin stained cross-section of breast cancer showing (A) high grade IDC with highly pleomorphic tumour cells and mitotic figures; (B) extensive ILC with concentric stromal infiltration by tumour cells around a duct uninvolved with cancer. Ductal and lobular basements membranes cannot be demonstrated. Source: WebPathology: Visual survey of surgical pathology.



**Figure 1. 7. Human breast tumour subtypes and possible link to normal breast epithelial hierarchy.** Shown is the possible origin of breast cancer subtypes from specific parts of the human breast. Stem cells appear to give rise to the claudin-low group; luminal progenitor cells to basal subtype; while different stages of maturation leading to a fully mature ductal cell transform to the Luminal B, HER2 and Luminal A subgroups. Source: Visvader, 2009.

<b>Classification of breast cancer</b>	<b>Anatomical origin</b>	<b>Prevalence</b>
<b>Ductal carcinoma</b>	Lactiferous ducts	70 – 80%
<b>Lobular carcinoma</b>	Lobules	10%
<b>Inflammatory</b>	All parts of the breast	1% - 4%
<b>Paget's</b>	Nipple and areola region	1% - 2%
<b>Others</b>	<b>Classed on special features</b>	
<b>Tubular</b>	Not clear	Not known
<b>Medullary</b>	Not clear	3% – 4%
<b>Adenoid Cystic</b>	Not clear	0.1%
<b>Lymphoma</b>	Lymphoid tissue	<1%

**Table 1.1 Anatomical classification and prevalence of breast cancer subtypes (worldwide).** Modified from Dabbs, 2012.

<b>Breast cancer Classification</b>	<b>Hormone receptor profile</b>	<b>Clinical characteristics</b>	<b>Example of breast cancer cell line</b>
<b>Luminal A</b>	ER+ PR+ HER2-	Responsive to chemotherapy and hormonal manipulation; low Ki67 expression (< 14%)	MCF7 T47D
<b>Luminal B</b>	ER+ PR+ HER2+	Moderate chemotherapy response; high Ki67 expression (> 14%); response to trastuzumab	BT474
<b>Basal</b>	ER- PR- HER2-	Good chemotherapy and trastuzumab response; high ki67 expression (>14%); no response to hormones; EGFR overexpressed	MDA MB 468
<b>Claudin low</b>	ER- PR- HER2-	Moderate chemotherapy response; low Ki67 expression (<14%), claudin-3; no hormone response	MDA MB 231
<b>HER2</b>	ER- PR- HER2+	Good response to chemotherapy and trastuzumab; No response to hormones; high Ki67 expression (> 14%)	SKBR3

**Table 1.2. Classification of breast cancer based on gene expression profile.**

Modified from Prat *et al.*, 2010; Holliday and Speirs, 2011; Park *et al.*, 2012; Maciejczyk, 2013. Key: ER - estrogen receptor; PR – progesterone receptor; HER2 – human epidermal receptor 2.

Taken together, the importance of breast cancer classification lies in the identification of particular subtypes that could stratify patients into specific treatment groups (Park *et al.*, 2012). The stratification allows the delivery of specific therapies to particular types of breast cancers and is partly responsible for the significant improvement in survival relating to breast cancer over the past four decades (Barrett, 2010; Park *et al.*, 2012). The anatomic type and gene expression profile of a breast tumour further contribute in building a profile for a given patient that could define management strategies. There are other prognostic factors that are also considered in deciding treatment options in breast cancer.

### **1.3 Prognostic factors in breast cancer**

Prognostic factors for breast cancer can be classified into three broad categories based on their relative importance and frequency of application in clinical practice. These factors are shown in table 1.3 and include category 1 factors which have been clearly validated with good clinical evidence supporting their role in breast cancer prognostication; category 2 factors used by some groups in deciding treatment but not yet in general clinical practice while category 3 factors are still at various stages of research. With respect to category 1 factors, the 'tumour, node, metastasis' (TNM) class combines the size and extent of the tumour, involvement of axillary lymph glands and presence or absence of distant metastasis, to define the stage of disease and assigns a prognostic evaluation to it. Perhaps the earliest evidence for the utility of TNM as a prognostic factor came from a large scale study of breast cancer prognostic factors which involved 24,740 women drawn from the Surveillance, Epidemiology and End Result (SEER) Program of the National Institutes of Health of the United States of America (Carter *et al.*, 1989). This report showed that breast cancer survival ranged from 45.5% in patients with tumour size of 50mm or more with involvement of axillary lymph glands to 96.3% in those patients with tumours less than 20mm and no lymph node involvement. Indeed both the size of the tumour and nodes in this report, served



as independent prognostic factors whose effects were also additive. Further work identified a subset of node negative patients based on tumour type and size with relapse-free survival rate of 91% at 10 years and 87% at 20 years (Rosen *et al.*, 1993). The authors of this latter report suggested that these patients might not need adjuvant therapy and reinforced the utility of TNM staging as powerful prognosticators of outcome in breast cancer. The TNM classification has been updated (Sobin *et al.*, 2009) to include the effect of such factors as lymph gland micrometastases (<2.0 mm) which were detected in 15.9% of 3887 patients previously assigned as node negative (Weaver *et al.*, 2011). The presence of micrometastasis in the latter study was found to be an independent prognostic factor in those patients initially thought to be node negative. Other category 1 prognostic factors such as hormone receptor (HR) status has been shown to possess predictive power for response to therapeutic and adjuvant hormonal treatment while mitotic count, a major component of histologic grading, could indicate how aggressive a tumour is, as the less differentiated it is, the more aggressive and therefore of worse prognosis (Fitzgibbons *et al.*, 2000; Tang and Gui, 2012). Another category 1 prognostic factor also in routine clinical use is the human epidermal receptor (HER) 2 status. Amplification of the *HER2* gene is present in about 20% - 30% of breast cancers and is associated with overexpression of p185 – a transmembrane protein – whose expression leads to reduced survival but greater response to doxorubicin chemotherapy and trastuzumab, a monoclonal antibody (Fitzgibbons *et al.*, 2000; Tang and Gui, 2012). On account of the improved response to these agents, acquisition of HER2 amplification has changed those tumours that harbour it from poor prognosis breast cancers to better prognosis breast cancers with improved survival.

Category 1	Category 2	Category 3
Tumour, node, metastasis (TNM)	Proliferation markers (Ki67)	DNA ploidy analysis
Grade	Lymphatic invasion	Microvessel density
Histologic type	P53	EGFR
Hormone receptor (HR) status		Transforming growth factor
Mitotic figure count		pS2, cathepsin
HER2 status		Multidrug resistance proteins
		Markers of tumour progression (CA15.3, uPA, PAI-1, CA 27.29)

**Table 1.3. Prognostic factors in breast cancer.** The categories define the relative importance and application of prognostic factors in clinical practice. Category 1 factors are the most frequently used in prognostication and deciding therapy while category 3 factors are still being validated. Adapted from Fitzgibbons *et al.*, 2000; Maciejczyk, 2013; Patani *et al.*, 2013.

Regarding category 2 factors, there is much research work in progress with regards to the impact of proliferation markers like Ki67 on breast cancer outcome and it is proposed that Ki67 expression levels may serve to identify potentially chemo-responsive patients (Wang *et al.*, 2011; Dowsett *et al.*, 2011). It is not clear however what the best cutoff point for Ki67 protein expression should be and as shown in table 1.2 it is far from certain that a certain Ki67 cutoff would determine chemotherapy response. The potential prognostic factors listed in category 3 are all subjects of intense research. Of note is the epidermal growth factor receptor (EGFR) whose suppression in the triple negative breast cancers (TNBC) that overexpress it, has been shown to render those cells more susceptible to cell death with the erlotinib - doxorubicin drug combination (Lee *et al.*, 2012). The authors of this latter report have therefore suggested that EGFR expression could be a prognostic factor in that breast cancer subtype. As more research findings emerge, the number of prognostic factors relevant to clinical management might increase and could become more tailored to specific breast cancer subgroups. Aside from the molecular factors discussed above, primary gene defects contribute to breast cancer onset and can also impact on the management strategy for a given patient.

#### **1.4 Genetic alterations in breast cancer**

Breast cancer can be familial (inherited) or sporadic (no known cause) and the primary mechanisms that drive the disease could be genetic, epigenetic, environmental or a combination of all three factors. Whatever the mechanism, the underlying genetic alteration in familial breast cancer is invariably mutations and the genes that undergo mutations can be divided into high penetrance and low penetrance genes (reviewed in Bogdanova and Dörk, 2012). Penetrance refers to the probability that a particular gene defect would lead to cancer formation in an age-specific manner when other competing causes are ruled out (Chen *et al.*, 2006). If the probability is high, then that gene defect is said to have high penetrance and vice versa. Of particular interest is the germ line affection of *BRCA 1* and *BRCA 2* genes which is seen in about 30% of familial cases but

present in less than 5% of the general population of breast cancer cases and represent the main high penetrance genes (Siegel *et al.*, 2013). *BRCA1* mutation is common (~80%) in women who have a family history of both breast and ovarian cancer but the incidence of this mutation falls to about 15-20% in women with a family history of only breast cancer (Chen and Parmigiani, 2007). *BRCA1* mutation carriers have a 60% - 80% (35% – 60% for *BRCA2*) lifetime risk of developing breast cancer with variable penetrance which however is highly dependent on the population studied (Chen and Parmigiani, 2007). Breast cancer patients that harbour the *BRCA* gene are usually younger at diagnosis (34.5 versus 37 years) and generally possess hormone receptor / HER2 negative and worse grade tumours compared to non-carriers (Nilsson *et al.*, 2014). The main mechanism responsible for the predisposition to cancerous malformation in this genetic condition resides in the loss of error-free homologous recombination in DNA double strand break (dsb) repair mediated by *BRCA 1* and *BRCA 2* proteins (Schlacher *et al.*, 2011). Cells that lack these proteins are forced to repair DNA via nonhomologous end joining, an error – prone process that leads to genomic instability. The importance of identifying gene defects in breast cancer with respect to patient management lies in the potential for personalised treatment. In patients with *BRCA* gene mutations, the defect in double strand break repair mechanisms was found to render breast tumour cells particularly susceptible to destruction by poly ADP-ribose polymerase (PARP) inhibitors via the process of synthetic lethality (Bryant *et al.*, 2005). In this process, the combined effect of mutation in both *BRCA1* and *BRCA2* genes (rather than the effect of a mutation in only one of the genes) is lethal to cells (Bryant *et al.*, 2005). PARPs repair DNA nicks via base excision processes and on inhibition is trapped at sites of single strand DNA breaks so that repair cannot proceed. With subsequent cell division, the single strand break is converted to a double stranded break rendering the cell particularly sensitive to destruction (Tutt *et al.*, 2010). In these *BRCA* gene carriers, normal cells are relatively spared because the tumour cells invariably possess both the mutant allele and a deleted or mutated wild type within the tumour.

Also the cancer cells are presumably cycling faster and possess a higher DNA damage load. On account of its potential effect, much research is currently in progress clarifying the role of PARP inhibitors in human breast cancer among *BRCA* mutation carriers (Bryant *et al.*, 2005; Tutt *et al.*, 2010). Further impact of *BRCA* germ line defects with respect to patient management options resides with the utility of prophylactic mastectomy. This surgical treatment modality reduced the risk for breast cancer in *BRCA* gene carriers by more than 95% consequently reducing breast cancer-specific mortality by about 90% in this group of breast cancer patients (Rebbeck *et al.*, 2004; Domchek *et al.*, 2010). Regarding the low penetrance genes, there are suggestions that they could function by modifying the phenotypic expression of the high penetrance genes and include the androgen receptor (AR) and *HRAS1* genes which modify the phenotype of the *BRCA1* gene leading to variable expression of the high penetrance gene (reviewed in Bogdanova and Dörk, 2012).

### **1.5 Epigenetic alterations in breast cancer**

Epigenetics describes a complex network of heritable traits that do not involve a change in the DNA sequence itself but revolve around methylation patterns, covalent modification of histone, and nucleosome remodelling (Jovanovic *et al.*, 2010). Methylation in turn describes the addition of a methyl group to the cytosine residue of a CpG island, a modification that effectively leads to the silencing of that gene and is observed in phenomena like genetic imprinting (Li, 2002; Jovanovic *et al.*, 2010). When methylation occurs at a tumour suppressor gene (TSG) promoter the effect could be loss of cellular control at that point and induction of cancerous change (Recillas-Targa *et al.*, 2006; Minning *et al.*, 2014). The mechanisms behind cellular transformation here could relate to the induction of point mutations; activation of proto-oncogenes; and genomic instability resulting from an imbalance in general methylation profiles (De Smet *et al.*, 2004). Histone modification (including acetylation, phosphorylation, methylation, SUMOylation, ADP ribosylation, deamination, ubiquitination) has an impact on

epigenetic regulation via the maintenance or alteration of binding sites for several factors involved in the DNA machinery (Esteller, 2007; Jovanovic *et al.*, 2010). Active chromatin is characterised by a histone hyperacetylation status and the activity of deacetylases therefore leads to heterochromatin formation and effectively gene silencing (Jovanovic *et al.*, 2010). Such changes in the histone code can lead to dysregulation and tumorigenesis since those interactions are necessary for DNA repair and cell division (Esteller, 2007; Jovanovic *et al.*, 2010). As mentioned previously, changes in methylation status (hyper- or hypo-) can also increase the risk of tumorigenesis via dysregulation of imprinting processes (Reik *et al.*, 2004; Girardot *et al.*, 2012). The mechanisms that contribute towards the increased risk of tumorigenesis on dysregulation of imprinting processes include unregulated cell proliferation from augmentation of gene expression and aberrant expression of those imprinted sites with resultant deletion of a monoallelically active gene (Reik *et al.*, 2004; Recillas-Targa, 2006). With respect to clinical utility, the methylation pattern of some cell cycle regulatory genes such as CDKN2A / p16 and CCNA1 has been found to be potential predictive markers for response to the anthracycline / mitomycin chemotherapy combination in some breast cancer patients (Klajik *et al.*, 2014). Regarding acetylation modification, the effect of histone deacetylation could be reversed with histone deacetylase inhibitors (HDACi) like vorinostat which has shown promise in combination therapy in a phase 1 / 2 trial in breast cancer patients (Ramaswamy *et al.*, 2012). Other important factors that could augment the risk for cancer development in this setting include the possible associated lack of chromatin components like the CCCTC-binding factor (CTCF), a deficiency that results in altered chromatin structure and insulator activity (Reik *et al.*, 2004; Recillas-Targa, 2006; Girardot *et al.*, 2012). The CCCTC-binding factor is thought to work via epigenetic mechanisms and an increasing body of evidence suggests that it is involved in the breast cancer phenotype. The following sections focus on this important protein and its involvement in breast cancer.

## 1.6 The CCCTC-binding factor (CTCF)

### 1.6.1 History, discovery and conservation of CTCF

In the search for nuclear factors involved in regulating the *c-myc* oncogene, a specific region about 200 base pairs (bp) upstream to the transcription start site (TSS) of this gene, was found to harbour two proteins (Lobanenkov *et al.*, 1990; Eilers and Eisenman, 2008). The first protein was similar to Sp1 (a transcription factor involved in cell growth, apoptosis and differentiation) and the second protein was bound to repeats of CCCTC in a regular fashion and was named CTCF (Lobanenkov *et al.*, 1990). The authors further discovered that removal of a nucleic acid sequence that bound both proteins (Sp1 and CTCF) led to a multi-fold rise in transcription of a stably transfected *c-myc* oncogene in chicken embryonic fibroblasts. They therefore suggested that CTCF was among other proteins probably involved in gene regulation at the chicken *c-myc* gene, specifically gene repression (Lobanenkov *et al.*, 1990). In further exploring the involvement of CTCF in transcription, Klenova *et al.* (1993) spontaneously induced specific mutations in three out of fifteen purines involved in CTCF binding using a polymerase chain reaction mediated site-directed mutagenesis procedure. In so doing they eliminated CTCF binding while other nuclear factors remained normally bound. On introducing a *c-myc* sequence bound to a reporter CAT gene, they discovered a 10 - and 3 - fold decrease in transcription in stably transfected erythroid and myeloid chicken cell lines respectively, suggesting a CTCF activator role. This possible function of CTCF as a gene activator is further supported by evidence from work on the mouse *H19* and *Tsix* genes as well as in the human *HLA-DRB1* and *HLA-DQA1* genes (Engel *et al.*, 2006; Donohoe *et al.*, 2007; Majumder *et al.*, 2008). The same characteristic binding to repeats of CCCTC was further discovered downstream of two conserved alternative transcription start sites (TSS) in the human / mouse *c-myc* gene suggesting among other things that CTCF was ubiquitously expressed in the genome (Filippova *et al.*, 1996). These findings raised speculation that CTCF could be important in cell replication.

The importance of this observation was magnified by evidence suggesting that a reduction in the CTCF gene product before fertilization prevented the development of the blastocyst and complete absence of CTCF gene expression was linked with death in the early embryonic phase in mouse (Fedoriw *et al.*, 2004; Splinter *et al.*, 2006). Moreover, evidence in a mammalian leukaemia cell line confirmed that when CTCF is overexpressed or knocked down then there are measurable effects that lead to changes in cell proliferative activities (Torrano *et al.*, 2005). There is further evidence linking oocyte transcription anomalies and problems with cycling of T lymphocytes in the thymus on account of CTCF depletion (Heath *et al.*, 2008; Wan *et al.*, 2008). Taken together it would appear that changes in CTCF gene expression affected multiple aspects of cellular activity. The structure and function of CTCF is well conserved across species from invertebrates like *Drosophila* to humans (Moon *et al.*, 2005).

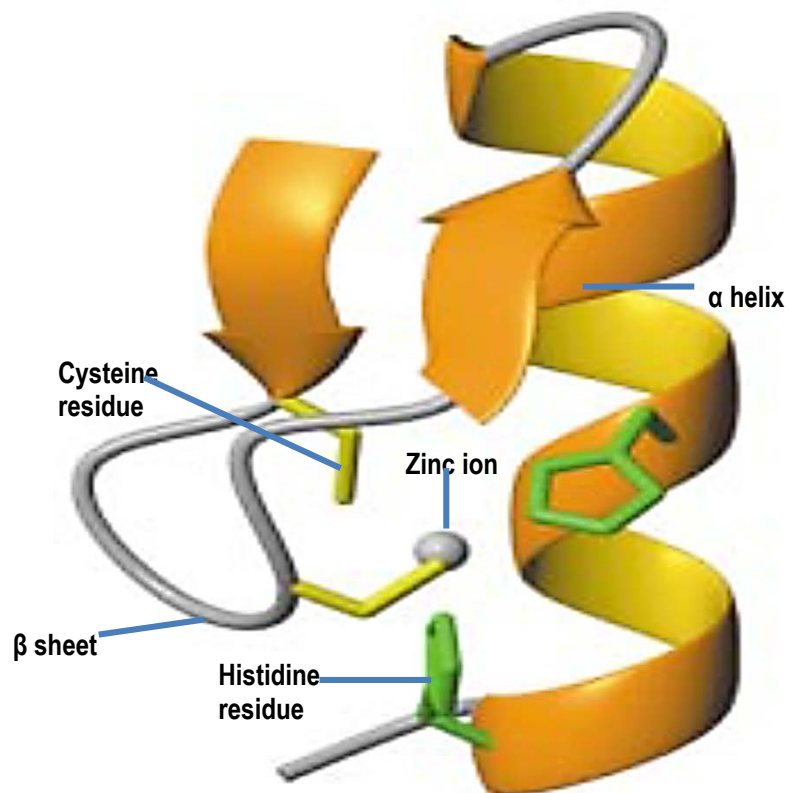
### **1.6.2 CTCF structure**

Structurally, CTCF is made up of three domains (C terminus, N terminus and the zinc binding domain) as shown in figure 1.8 and composed of a total of 727 amino acids with the DNA binding domain comprising 11 zinc finger (ZF) units (Klenova *et al.*, 1993). The zinc finger units shown in figure 1.9 incorporate zinc ion into a complex of cysteine and histidine residues and are classified into fold groups depending on the shape of the protein backbone (Razin *et al.*, 2012). The Cys<sub>2</sub>His<sub>2</sub> fold group is the best characterised and possesses a beta-beta-alpha fold helix configuration (Razin *et al.*, 2012). The alpha helix of the ZF is the recognition sequence for DNA binding sites and each recognition sequence can bind in a sequence-specific pattern to four or more DNA bases. Binding of a zinc finger can overlap with other zinc fingers and in this manner CTCF could make multiple simultaneous DNA contacts (Razin *et al.*, 2012). CTCF has 10 Cys<sub>2</sub>His<sub>2</sub> units and one Cis<sub>2</sub>HysCys unit (Klenova *et al.*, 1993). Only four of the 11 ZFs bind strongly to DNA and have been identified as ZF 4 to 8, the 8<sup>th</sup> becoming more important in the absence of the fourth and thought to have more non-sequence specific activity



compared to ZF 4 to 7 (Renda *et al.*, 2007). Flanking the ZF domain are the amino (N) - and carboxyl (C) - terminal domains (figure 1.7). These are polypeptide chains 265 and 148 residues in length respectively and are monomeric, unordered and without domains (autonomously folding units of stable secondary structure) (Martinez and Miranda, 2010). This flanking region of CTCF has been shown to interact with histones, the large subunit of polymerase 2 and other proteins involved in DNA interactions (Martinez and Miranda, 2010). The N terminal end of the ZF middle portion is the transcriptional start (Renda *et al.*, 2007) and there are suggestions that CTCF based on this structure might also function as a scaffold or adaptor that facilitates DNA interactions (Martinez and Miranda, 2010).





**Figure 1.9. Schematic structure of the classic Cys2 His2 zinc finger (ZF).** Showing  $\alpha$  helix,  $\beta$  sheet, and the cysteine and histidine residues in relation to zinc ion. The zinc ion offers stability to the zinc finger unit. Adapted from Laity *et al.*, 2001; Renda *et al.*, 2008.

### **1.6.3 CTCF functions**

CTCF has been implicated in an array of important functions such as gene regulation, intra and inter-chromosomal looping, imprinting, insulation and X chromosome inactivation. It is thought that the many ZF domains are capable of simultaneously binding multiple DNA sequences and that the resultant varied conformational states might be responsible in a large part for its involvement in extensive cellular activities. These activities are described below.

#### **1.6.3.1 Genome organiser and long-range chromatin interaction role**

The understanding of gene function and regulation is changing. The extensive review of Phillips and Corces (2009), noted that up until recently a linear model involving a straightforward sequence of events between enhancers, promoters and enzymes on one chromosome led to gene expression. There is evidence now suggesting that the location of chromosomes in the nucleus is not static with further evidence favouring the presence of transcription pits, which are factories containing RNA polymerase II, enhancers, locus control elements, and promoters necessary for gene transcription (Osborne *et al.*, 2004; Lanctôt *et al.*, 2007). CTCF has been found to mediate chromatin interaction activity at the *H19 / Igf2* gene locus, an interaction that placed regions of the chromosome naturally far apart, in close proximity (Murrell *et al.*, 2004). It is thought that the mechanism behind this long-range interaction could revolve around the ability of the C terminus of CTCF and its ZF domain to induce conformations that allowed looping out of the intervening DNA (Ling *et al.*, 2006). This activity could therefore bring relatively far flung cellular factors to transcription pits and mediate gene activity. In order to verify whether the CTCF long range chromatin interaction activity was on a genome-wide scale, chromatin immunoprecipitation experiments followed by paired end tag extraction and sequencing (CHIA-PET) were performed on mouse embryonic stem (ES) cells (Handoko *et al.*, 2011). The paired end tag extraction and sequencing is a chromosome

conformation capture (3C) (which studies spatial organisation of long genomic regions) – based assay that permits sequencing of interactions associated with an immunoprecipitated factor. The results of the CHIA-PET assay analysed in the report of Handoko *et al.* (2011) compared quite favourably with the results of traditional chromatin immunoprecipitation – sequencing (ChIP - Seq) tests performed by the same authors but the CHIA-PET assay needs verification by other authors to assess reproducibility and therefore reliability. The Handoko *et al.* (2011) report identified CTCF binding sites (PETs with two ends within 10kb on a single chromosome), intrachromosomal interacting PETS (where the two ends are on the same chromosome but more than 10kb apart) and interchromosomal interacting PETS (where the PET tags were mapped to different chromosomes). The CTCF interactome map created by the forementioned PETs contained multiple loops that encompassed different genes, connected multiple promoters and also specifically confirmed the *Igf2* DMR – *H19* ICR interaction locus as previously suggested by other authors (Murrell *et al.*, 2004). In effect, Handoko *et al.* (2011) confirmed that CTCF mediated long-range interactions and emphasised that enhancers, promoters and their cognate genes did not have to be in close proximity to work together. Using non - supervised clustering algorithms, they further discovered that CTCF interactions divided the genome into five domains which were defined by CTCF loops and based on specific histone modification patterns as shown in table 1.4. In effect, the domain in which the CTCF interaction occurred would determine whether that interaction would be activating (domain 1) or repressive (domain 2). Not all domains however possessed clear activating or repressive features and more work needs to be done to clarify this model of genome partitioning. Furthermore, what decides the partitioning of some interactions into one domain or the other is by no means clear but could speculatively relate to the effect of posttranslational modifications and / or protein partners.

	Frequency	Activating histone marks	Repressive histone marks	Miscellaneous features
<b>Domain 1 (activating)</b>	12%	Enriched in loops (H3K4me1 and me2 and H3K36me3)	Depleted  K9, K20, K27 methylation marks	
<b>Domain 2 (repressing)</b>	11%	Under- represented	Extensively represented	
<b>Domain 3</b>	19%	1 and 2 within loop regions; 3 at boundaries; H3K36me3 outside of loops		
<b>Domain 4</b>	31%	At one side of loop	At opposite site of loop	Possibly domain barriers
<b>Domain 5</b>	27%			No specific pattern of histone modification seen

**Table 1.4. Categories of CTCF chromatin loop domains.** Based on the clustering of histone modifications (unique epigenetic states) maintaining the 3-dimensional structure of chromatin. The domains generally partition the genome into unique activating or inhibiting regions. Adapted from Handoko *et al.*, 2011.

### **1.6.3.2 CTCF regulation of gene expression**

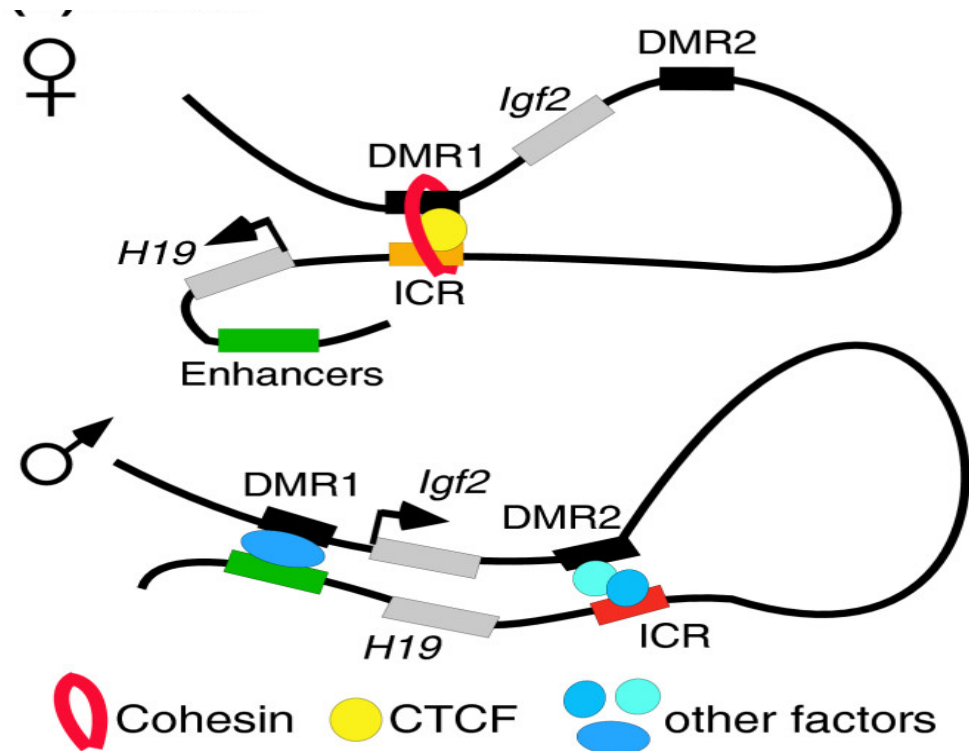
While CTCF may function as a gene repressor or gene activator (Lobanenkov *et al.*, 1990; Klenova *et al.*, 1993; Filippova *et al.*, 1996) it is not clear whether these activities are a result of CTCF acting as a transcriptional factor (TF). What is clear however, judging by activity at the major histocompatibility complex class II (MHCII) locus is that CTCF binding and looping between upstream sequences occurred before transcriptional activation at that gene locus (Majumder and Boss, 2010). In effect, via long-range interactions, CTCF is able to bring gene promoters and enhancers together and mediate transcriptional activity. Furthermore, evidence from 3C-Seq confirmed the involvement of CTCF in genome wide long range interactions in limiting interactions at the immunoglobulin K (IgK) enhancers with resultant control of gene expression at that locus (de Almeida *et al.*, 2011). There are suggestions also that these processes might be responsible for CTCF effect on transcriptional regulation at the protocadherin- $\alpha$  cluster of genes and the angiogenin (ANG) and ribonuclease 4 (RNASE4) superfamily (Hirayama *et al.*, 2012; Sheng *et al.*, 2014). Disruptions of these genes have been shown to lead to problems with neuronal development, neurodegenerative disorders and some cancers such as breast and prostate cancer (Hirayama *et al.*, 2012; Sheng *et al.*, 2014). It would appear therefore that CTCF-induced DNA loop formation might be part of the basis for the involvement of CTCF in gene activation or repression. The result of that involvement might depend on the context of the tissue-specific chromatin organization (activation or repression) as suggested by Handoko *et al.* (2011).

### **1.6.3.3 CTCF insulation and genomic imprinting role**

Insulation refers to the ability to prevent interaction between two adjacent gene regions in a position-dependent manner (Ishii and Laemmli, 2003). Insulators therefore are DNA sequences that act to maintain the physical integrity of transcriptional domains by inhibiting interference from repressive heterochromatin (barrier) or activating signals

emanating from gene enhancers to promoters (enhancer blocking) (Ishii and Laemmli, 2003). The ability to effect insulation appears also to revolve round DNA looping processes (Kurukuti *et al.*, 2006; Zhang *et al.*, 2011). In order to establish the involvement of CTCF in insulation, a genome-wide study using ChIP-Seq found a small but statistically significant proportion of CTCF-binding sites concentrated at boundaries between active and repressive domains marked by histone H2A lysine 5 acetylation (H2AK5Ac) and histone H3 lysine 27 trimethylation (H3K27me3), respectively (Cuddapah *et al.*, 2009). These authors reported that the CTCF binding sites were mostly intergenic in location and though they found considerable overlap between binding sites in the three cell types studied (HeLa, CD4+ and Jurkat), there was minimum overlap between those CTCF marked domains in HeLa and CD4+ cells. While confirming the importance of CTCF in barrier function, they suggested that the link between CTCF and domain boundaries could be cell type- specific. The most important relevance of CTCF as an insulator however may relate to imprinting wherein certain genes are expressed in a parent-of-origin-specific manner (mono-allelic expression) (Singh *et al.*, 2012). The best example for this phenomenon in vertebrate cells is probably at the mouse imprinted *Igf2* - *H19* locus (Fedoriw *et al.*, 2004; Wallace and Felsenfeld, 2007). This site, shown in figure 1.10, has the imprint control region (ICR) lying between the *Igf2* gene and downstream enhancers. The *H19* gene locus is also downstream to the ICR. The ICR possesses four CTCF binding sites and is unmethylated in the maternally derived allele (Wallace and Felsenfeld, 2007; Singh *et al.*, 2012). CTCF is therefore able to bind this region and loop out the intervening DNA in such a manner that the DNA methylated region 1 (DMR 1) which is upstream to the *Igf2* gene comes into contact with the ICR as shown in figure 1.10. In that way it blocks the transcription of the *Igf2* gene which is now held in a loop while at the same time allowing expression of the *H19* gene. The opposite happens in the paternally derived allele where due to methylation at the ICR, CTCF cannot bind therefore downstream





**Figure 1.10. CTCF and imprinting at the *Igf2* / *H19* gene locus via loop formation.**

Monoallelic expression at the *Igf2*-*H19* locus is regulated by binding of CTCF to the imprinted control region (ICR). On the maternal allele, CTCF mediates interactions between ICR and DNA methylated region 1 (DMR1) that also involves joining of the DNA strands by cohesin, insulating *Igf2* from the influence of downstream enhancers. Methylated ICR sequences prevent CTCF from binding to the ICR on the paternal allele, allowing downstream enhancers to switch on *Igf2* transcription. Source: Ong and Corces, 2009.

enhancers are able to switch on transcription of the *Igf2* gene while inhibiting the *H19* gene. In effect the *Igf2* gene is paternally expressed while the *H19* gene is maternally expressed. Strict control of insulation at this imprint site appears to be essential for normal development as mice that acquired mutations at the CTCF ICR or had biallelic expression at this site developed phenotypic abnormalities and humans could develop clinical syndromes like the Silver – Russell syndrome characterised by dwarfism (Szabó *et al.*, 2004; Singh *et al.*, 2012).

#### **1.6.3.4 X Chromosome inactivation**

Further support for the role of CTCF in chromosomal organisation comes from its involvement in X chromosome inactivation. The female possesses two X chromosomes but one of them is silenced to equalise the contents of this chromosome between males and females. This entirely random process (X chromosome inactivation) involves counting, choice and mutually exclusive silencing and is coordinated from the X inactivation centre (Xic) which has been shown to have many binding sites for CTCF (Xu *et al.*, 2006; Xu *et al.*, 2007; Tsai *et al.*, 2008). The exact mechanism governing this whole process is still debated. However the involvement of CTCF is clear as CTCF knockout was associated with deregulation of the process which required homologous X chromosome pairing and associated intra- and inter-chromosomal loop formation (Donohoe *et al.*, 2007; Xu *et al.*, 2007).

In summary, it would appear that a major impact of CTCF on cellular function revolved around its ability to organise chromatin into domains of activation or inhibition. Also on account of its ability to reach out (via loop formation) across considerable distances it is able to bring nuclear factors together and in so doing determine the nature of nuclear processes like transcription. From the literature, it is also clear that the effect of CTCF could be completely opposite (as with transcription activation versus repression) suggesting that other factors like CTCF binding sites, post translational modifications

and protein partners might presumably play a role in the resultant effect of CTCF binding (Phillips and Corces, 2009; Zlatanova and Caiafa, 2009).

## **1.7 CTCF DNA – binding sites**

The functional effect of protein binding might depend on the characteristic of the DNA binding site itself. It has been suggested that the majority of vertebrate transcription factor–binding sites were better divided up into multiple classes since prediction of binding could be better made based on multiple DNA binding classes rather than a single all-encompassing DNA binding sequence (Hannenhalli and Wang, 2005). The section below describes CTCF DNA binding sites and their impact on CTCF action.

### **1.7.1 Distribution and characteristics of CTCF DNA - binding sites**

In a bid to explore the association between CTCF and insulation in the vertebrate genome, an extensive effort using ChIP assays to localise CTCF DNA binding sites, followed up with genome tiling micro arrays, confirmed 13804 sites in the human IMR90 fibroblasts (Kim *et al.*, 2007). Using a discriminatory motif enumerator, a bioinformatics tool that applied position weight matrices, the authors also confirmed the presence of a consistent 20 base pair (bp) DNA - sequence motif for three quarters of identified CTCF binding sites. The remaining CTCF DNA binding sites did not fit into this consensus sequence suggesting that CTCF was capable of other DNA interactions and possibly implicated the effect of multiple zinc finger binding combinations. The ChIP and high throughput DNA sequencing (ChIP-Seq) method is recognised to be of particular utility in identifying DNA-protein interactions and was used to detect 20,262 CTCF binding sites in human CD4<sup>+</sup> T cells; 39,609 sites in the mouse embryonic stem (ES) cells; 19308 sites in HeLa and 19,572 in Jurkat cells (Chen *et al.*, 2008; Cuddapah *et al.*, 2009). In effect, there appeared to be differences in the number of CTCF binding sites in different cell types and it would not be surprising that the pluripotent undifferentiated embryonic stem (ES) cells would possess a greater number of binding sites. While the

great number of binding sites churned out by ChIP-Seq is impressive, it is not clear whether they are all of biologic significance. This experimental method first immunoprecipitates a protein-DNA fragment using specific antibody to the protein in question, then subjects the captured DNA fragment to sequencing and computational analysis (Bailey *et al.*, 2013). The analysis that involves extensive bioinformatics is fraught with pitfalls as reproducibility, cross-correlation, and metrics quality cut-offs determine the significance of identified sequences (Bailey *et al.*, 2013). On account of the infancy of this method at the time of the reports of Kim *et al.* (2007), Chen *et al.* (2008) and Cuddapah *et al.* (2009), it is not clear whether the rigorous standards of Bailey *et al.* (2013) were followed. The ChIP-Seq nevertheless remains the method of choice for unravelling genome-wide protein – DNA interaction and in IMR90 fibroblasts (derived from fetal lung with characteristics of smooth muscle), the location of CTCF binding sites was reported as 46% intergenic, 22% intronic, 12% exonic, and 20% within 2.5 kb of promoters (Kim *et al.*, 2007). More detailed computational analysis of the data from the human CD4+ T cells employing a more sophisticated algorithm which possibly had a higher sensitivity and specificity showed a total of 26,814 binding sites with 45% intergenic, 7% 5'UTR, 3% exonic, 29% intronic, 2% 3'UTR, and 13% of sites within 5kb of the transcription start site (Jothi *et al.*, 2008). The majority of CTCF binding sites are therefore some distance from TSS lending support to the observation that its activity in transcriptional activity may be somewhat different from traditional transcriptional factors (see section 1.6.3.2). Also, the intergenic location of most CTCF binding sites supported the idea that it was involved in significant insulation activity confirming the initial hypothesis of Kim *et al.* (2007). Regarding the consensus DNA binding site sequence, there are two conserved cores in the 20 base pair (bp) motif: the first between the 4<sup>th</sup> and 8<sup>th</sup> bases inclusive and the second between the 10<sup>th</sup> and 18<sup>th</sup> bases (Essien *et al.*, 2009). Further work on the consensus DNA binding site narrowed the most important area for binding to a 12bp core site bound with high affinity by CTCF ZF 4 to 8, ZF 4 interacting with the 3' end and ZF 8 with the 5' end of DNA (Renda *et al.*, 2007).

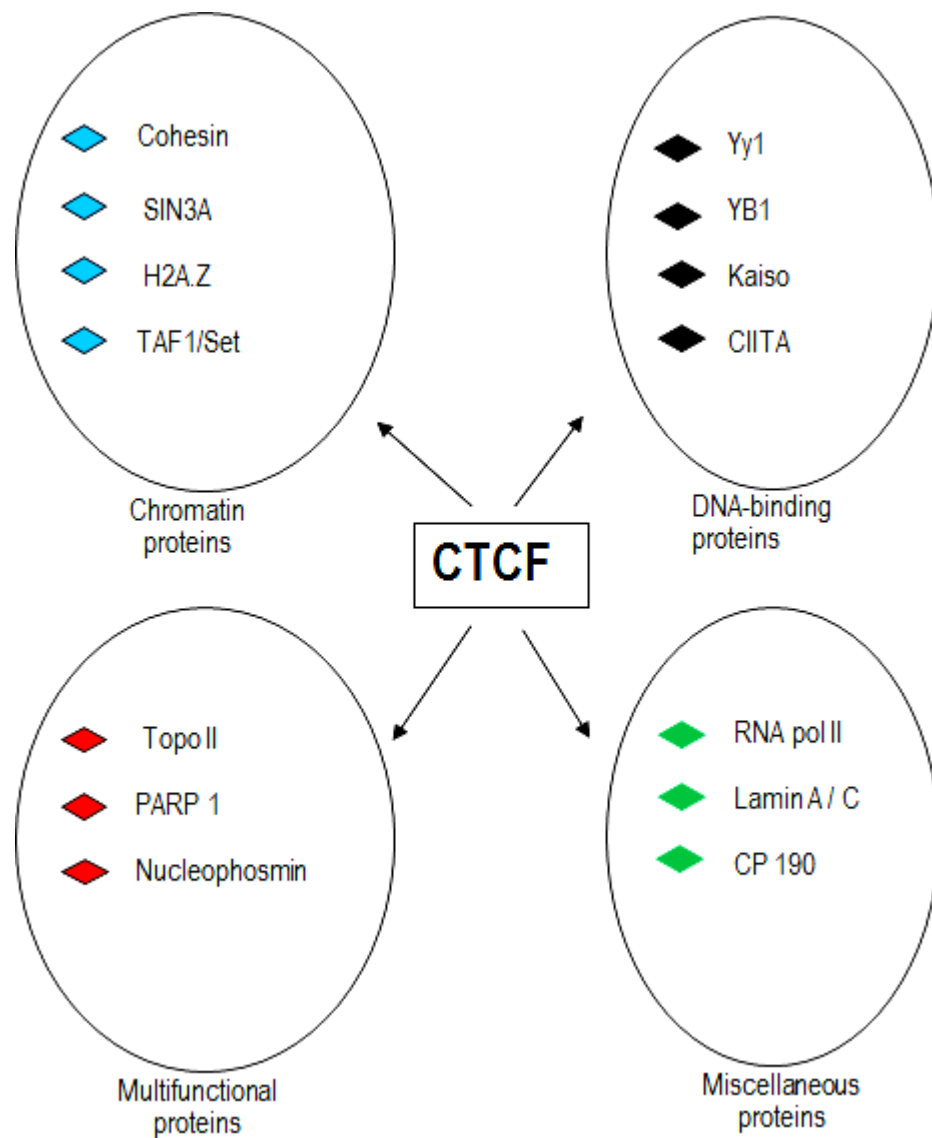
### 1.7.2 Classification and functional effect of CTCF binding sites

On account of the extensive distribution and involvement of CTCF in cellular processes and its potential as a transcriptional factor (TF), Essien *et al.* (2009) investigated whether the TF DNA binding site classification of Hannonhalli and Wang (2005) was also applicable to CTCF. They employed positional weight matrix (PWM) indices to represent the CTCF motif and used a scan tool to get the best possible match between the sites studied and the consensus binding site in the over 26000 CTCF binding sites identified in human CD4+, HeLa, Jurkat and IMR90 cells. Based on their results they confirmed that CTCF binding sites were grouped into low occupancy, medium occupancy, and high occupancy reflecting an increasing degree of homology between sites screened and the consensus site (Essien *et al.*, 2009). They showed that this classification translated into different functional effects: low occupancy sites were more cell-type specific, had a greater concentration of active histone marks (H3K27me1) and in effect behaved like euchromatin, and had less evolutionary change between mouse and human. High occupancy sites on the other hand were more associated with gene co-expression and repressive histone marks (H3K27me2 and H3K27me3) or heterochromatin. While PWM is used extensively in bioinformatics to characterise motifs in nucleotide sequences, there are few statistical tests available to evaluate the significance of PWM output and results based on this method may be subject to dispute (Xia, 2012). Though it is not clear how accurately the PWM predicted the CTCF motif, the report of Essien *et al.* (2009) served to reinforce the previously described activity of CTCF in chromatin organisation (Kim *et al.*, 2007; Handoko *et al.*, 2011) and that transcription factor binding sites could have multiple classes (Hannonhalli and Wang, 2005). CTCF target sites would therefore appear to be a heterogeneous group of DNA sequences. The biological significance of this heterogeneity is not yet clear however alterations in the DNA binding sequence could affect CTCF activity as single nucleotide mutations at CTCF binding sites can result in completely different CTCF binding affinity

and effects. For instance, CTCF binding could be reversed by a mutation in a single nucleotide contact region (outside of the consensus motif) in the *myc-A* CTCF binding site (Payer and Lee, 2008). Furthermore non-experimental single nucleotide mutations like the C-43-A mutation in the *Xist* gene abrogates CTCF binding while C-43-G mutation does the complete reverse (Pugacheva *et al.*, 2005). It might therefore be surmised that alterations at these sites could render cells more susceptible to malignant transformation. It would therefore appear that a complicated interaction between the primary CTCF binding sequence, the resultant ZF combination and the context within which the particular CTCF binding site functioned would determine the functional effect of CTCF binding (Ohlsson *et al.*, 2010). Though about a quarter of CTCF binding sites detected *in vivo* do not conform to the 20bp consensus motif shown by Kim *et al.* (2007), CTCF is nevertheless able to bind these DNA sequences suggesting either a direct effect or a possible association with protein partners.

## **1.8 CTCF protein partners**

The functional effect of CTCF - DNA interaction could be modified by protein partners. It is not clear however whether CTCF recruits a protein to interact with or whether it facilitates the action of a native protein at a given location (El-Kady and Klenova, 2005). According to the extensive review of Zlatanova and Caiafa (2009), CTCF protein partners can be broadly divided into four groups namely chromatin proteins, DNA binding proteins, multifunctional proteins, and a miscellaneous group without any particular distinguishing group feature. The different CTCF protein partner groups are shown in figure 1.11 and individual protein partners with possible links to cancerous malformation (especially breast cancer) are discussed below.



**Figure 1.11. Schematic representation of classes of CTCF protein partners.** Key: Top II- topoisomerase II; RNA pol II – RNA polymerase II; PARP 1 – poly ADP ribose polymerase 1; Yy1 – Yin and Yang 1. Source: Zlatanova and Caiafa, 2009

### 1.8.1 Chromatin CTCF - protein partners

Chromatin CTCF protein partners (as shown in Figure 1.11) include cohesin, a protein involved in chromosome segregation during mitosis and necessary for homologous recombination-dependent DNA repair (Wendt *et al.*, 2008). It has been shown via chromatin immunoprecipitation that 70% of identified CTCF binding sites are co-occupied by both CTCF and cohesin and both proteins are involved in regulating transcription at the *Igf2 / H19* imprinted site (Parelho *et al.*, 2008). The exact relationship between these protein interactions is still unclear because while cohesin interferes with CTCF, the latter has no effect with regards to the activity of cohesin during mitosis (Parelho *et al.*, 2008; Wendt *et al.*, 2008). This observation could be in keeping with a proposal identifying two different pools of cohesin activity, the first an immobile fraction bound to chromatin and needed for cellular cohesion and then a more mobile fraction that might mediate interaction with DNA binding proteins like CTCF (Zlatanova and Caiafa, 2009). With regard to cellular transformation and tumorigenesis it is possible that on account of the need for cohesin in homologous-recombination-dependent DNA repair and much like the action of the *BRCA* genes, disruption of the CTCF - cohesin interaction could impair cellular ability to repair DNA damage and therefore render them susceptible to cancerous malformation.

### 1.8.2 DNA-binding CTCF protein partners

Some of the DNA binding proteins that partner with CTCF are shown in figure 1.11. Kaiso, a zinc finger protein, was found to bind CTCF bait through the CTCF C-domain in a yeast two-hybrid screen and co-immunoprecipitated CTCF with specific anti-Kaiso monoclonal antibodies (Defossez *et al.*, 2005). The involvement of Kaiso in breast cancer had previously not been described but a recent report described the interrogation of Kaiso expression and linked this expression to molecular subtypes and pathologic indices in human invasive breast cancer (Vermuelen *et al.*, 2012). This work found an



association between nucleus-bound Kaiso and aggressive breast cancer phenotype (high histologic grade, ER $\alpha$  negativity, HER-2 positivity and EGFR overexpression). The partnering of CTCF with Kaiso might therefore indicate a functional involvement of CTCF in breast tumorigenesis. Other DNA partner proteins include the Yin and Yang 1 transcription factor (Yy1), a 4-zinc finger factor necessary for embryogenesis but also involved in X-chromosome inactivation mediated by CTCF (Gordon *et al.*, 2006). Both proteins (CTCF and Yy1) have many paired binding sites at the *Tsix* region of the X-chromosome inactivation centre and the Yy1 protein binds the N terminal of CTCF to trans activate *Tsix*, an activity that is thought to be more potent than either acting alone (Donohoe *et al.*, 2007). Yy1 has been shown to associate with cell cycle signalling pathways in ER+ breast cancers and that interaction contributed to G1-phase progression and proliferation activation in estrogen responsive breast cancers (Cicatiello *et al.*, 2004). The co-expression of CTCF and YY1 could therefore also reinforce the previously suggested involvement of CTCF in the breast cancer phenotype.

### **1.8.3 Multifunctional CTCF protein partners**

Within the multifunctional protein group in figure 1.11, PARP 1 and CTCF were discovered to form functional complexes through single and serial chromatin immunoprecipitation assays in mouse DNA (Farrar *et al.*, 2010). In this paper, PARP1 was overexpressed and it is not clear whether the effect of an artificially exaggerated quantity of PARP1 in the cell would be of biologic significance. PARP 1 is also involved with CTCF in gene imprinting and ribosomal gene expression (Yu *et al.*, 2004; Torrano *et al.*, 2006). With respect to cancer, PARP 1 plays a key role in cellular DNA repair processes and inhibition of this enzyme limits cell DNA repair and augments the susceptibility of cells that already possess repair defects as observed in *BRCA* deficient breast cancer cells, to destruction (Tutt *et al.*, 2010). This characteristic is currently being tested in Phase 1 and 2 human breast cancer trials and provides continuing evidence for the possible role of CTCF in breast cancer (Tutt *et al.*, 2010).

#### **1.8.4 Miscellaneous CTCF protein partners**

Among the miscellaneous group of protein partners (figure 1.11), the large subunit of the RNA polymerase II (RNA pol II) was found to interact directly with CTCF during serial chromatin immunoprecipitation assays in proliferating HD3 cells (Chernukhin *et al.*, 2007). RNA pol II is part of the core cellular transcription machinery and aberrations of Pol II and its multiple interactions at this site clearly could lead to malignant transformation. The complexing of CTCF with this protein would therefore suggest that CTCF integrity was needed to maintain normal cellular homeostatic mechanisms.

#### **1.8.5 Summary**

In summary, there appears to be a complicated network of proteins involved in CTCF action, made even more so by extensive protein-protein and protein–CTCF crosstalk. For instance, as noted by Zlatanova and Caiafa (2009), nucleophosmin interacts with PARP1, which is a known partner of Yy1, and both in turn interact with CTCF. Clearly more work needs to be done to tease out the strings that bind these proteins to themselves and to CTCF so as to among other things determine the relative contribution of CTCF binding site homology on the one hand and protein partners on the other to CTCF function at any given site.

### **1.9 Regulation of CTCF activity**

The extensive effects of CTCF in the cell are controlled by normal cellular mechanisms and include methylation and post translational modification (PTM) processes (El-Kady and Klenova, 2005; Wallace and Felsenfeld, 2007; MacPherson *et al.*, 2009; Witcher and Emerson, 2009). These control processes are important since a loss of balance between gene activator and inhibitor signals could lead to malignant transformation and are described below.

### **1.9.1 Methylation**

Methylation refers to the transfer of a methyl group to CpG dinucleotides, a process that effectively silences activity at that site (Fedoriw *et al.*, 2004; Wallace and Felsenfeld, 2007). Regulation of CTCF activity could be achieved with methylation occurring in and around CTCF binding sites. A previously mentioned important example is the *Igf2 / H19* gene locus where paternal-specific methylation of the imprint control region (ICR) determined monoallelic CTCF binding which results in maternal-specific allele gene expression (Fedoriw *et al.*, 2004; Wallace and Felsenfeld, 2007). The presence or absence of methylation at this locus then clearly determines CTCF binding and thereby regulates monoallelic gene expression at that site. As previously noted in section 1.6.3.3, a disruption of this regulatory process could lead to cell transformation. Further support for the effect of methylation on CTCF activity away from the *Igf2/H19* locus was provided using *BRCA1*-methylated breast cancer cell lines - UACC3199 and HCC38 (Xu *et al.*, 2009). This paper studied the effect of methylation at the *BRCA1* gene promoter and noted that CTCF binding to this promoter occurred only in the unmethylated state (unlike Sp1 whose binding was unaffected by methylation status). The authors therefore concluded that methylation processes regulated CTCF binding at this locus.

### **1.9.2 Post translational modifications (PTM)**

Posttranslational modification of CTCF is involved in CTCF regulation and could be a strong contributor to its activity. These processes which include poly (ADP-ribosyl) ation (PARylation), phosphorylation and SUMOylation serve to place certain boundaries on the CTCF protein and by so doing regulate the extent of its interactions.

#### **1.9.2.1 PARylation**

PARylation is the addition of poly-ADP ribose (PAR) groups to chromatin proteins, a process catalysed by poly-ADP ribose polymerases (PARPs) (Beneke, 2012). The PAR groups transfer negative charges to the proteins they bind and alter their interaction with

other proteins including DNA (Beneke, 2012). CTCF PARylation, lost on PARP inhibition, is associated with loss of insulation activities including gene imprinting at the *H19* ICR (Yu *et al.*, 2004; Jothi *et al.*, 2008) and also has effects on cell proliferation and tumorigenesis (Docquier *et al.*, 2009; Witcher and Emerson, 2009). These effects are discussed more extensively in chapter 3 of this thesis. Other functions of CTCF possibly regulated by PARylation include transcription of ribosomal RNA (Torrano *et al.*, 2006). This latter report showed that transfection of UR61 cells (a cell line derived from pheochromocytoma of the rat adrenal medulla) with GFP-CTCF led to cessation of nucleolar transcription compared to controls as detected by 5'fluorouridine (5'-FU) incorporation assays. Conversely, treatment of these transfected cells with 3-aminobenzamide (a PARP inhibitor) led to the restoration of 5-FU incorporation suggesting that CTCF PARylation directly controlled ribosomal RNA transcription and therefore regulated CTCF activity.

#### **1.9.2.2 SUMOylation**

The process of SUMOylation involves the covalent addition of a Small Ubiquitin-like Modifier (SUMO) to a protein factor. This binding alters protein stability, nuclear-cytoplasmic transport, cell cycle progression, transcriptional inhibition, and protein interaction with binding partners (Gill, 2005; MacPherson *et al.*, 2009; Yang and Chiang, 2013). Less than 10% of CTCF protein is SUMOylated however this modification of CTCF has been shown to lead to transcriptional repression at the *c-myc* P2 promoter (MacPherson *et al.*, 2009). Furthermore, CTCF-induced chromatin opening is prevented by SUMO-3 modification suggesting that this modification could be involved in insulation activities (Kitchen and Schoenherr, 2010). In effect, SUMO modification regulates CTCF processes such as transcription and insulation and aberration of these processes could present a template for cellular transformation.

### **1.9.2.3 Phosphorylation**

An extensive review of protein phosphorylation noted that phosphate groups associated with transcription factors (TFs) by regulating the amount of time the factors spent in the nucleus; initiated TF degradation; and interfered with TF binding to DNA by changing chromatin structure (Abrantes *et al.*, 2014). Specific phosphorylation of CTCF has been mapped to four serine residues in the C terminal region namely positions 604, 609, 610 and 612 with the critical site at the 612 residue (Klenova *et al.*, 2001). It has been shown that phosphorylation at these residues and especially at the 612 residue converted CTCF from gene repressor to a gene activator suggesting that CTCF was associated with both transcriptional repression and activation depending on the presence or absence of phosphorylated residues (El-Kady and Klenova, 2005). It is possible that a switch of CTCF depending on the context could lead to an imbalance in cellular function that could be tumour inducing. The functional effect of this modification with respect to breast cancer however is yet to be elucidated.

## **1.10 CTCF and tumour suppression**

CTCF could be a gene activator (usually after modification) as alluded to in the previous section but its repressor activity with respect to carcinogenesis appears to predominate. The mechanisms described in this section could serve to explain the possible general action of CTCF in suppressing tumour formation and some of the evidence for this view is described below.

### **1.10.1 CTCF induces transcriptional repression of hTERT in tumours**

Tumours are characterised by the ability to grow indefinitely partly because some tumours are able to generate telomerase, an enzyme involved in maintaining telomere length enabling those tumours to sustain high proliferation rates and bypass senescence

(Elenbaas *et al.*, 2001). In normal cells however, telomerase is expressed at low levels and with each division, the length of the telomere decreases up to a point that division is no longer possible and cells enter into senescence (Masutomi *et al.*, 2003). Telomerase is made up of an RNA component and the human telomerase reverse transcriptase (hTERT) which is expressed by only TERT positive cells and whose production is controlled by positive (for instance, estrogen receptor) and negative (for instance, Ap1 and p53) regulators (Mason *et al.*, 2011; Gómez *et al.*, 2013). It has been shown that CTCF binding inhibited hTERT transcription in cancer cell lines (Choi *et al.*, 2010; Meeran *et al.*, 2010). To prove that this effect of CTCF on hTERT occurred via methylation, HCT116 and breast cancer cell lines were treated with trichostatin and sulforaphane, agents that inhibit histone deacetylase (HDAC) and DNA methyltransferase 1 (Choi *et al.*, 2010; Meeran *et al.*, 2010). With the resultant demethylation at the hTERT promoter, CTCF was bound to hTERT and induced telomerase repression (Choi *et al.*, 2010; Meeran *et al.*, 2010). It is thought that this ability to suppress telomerase via epigenetic mechanisms might be one of the mechanisms through which CTCF played a role in tumour suppression in cancer including breast cancer.

#### **1.10.2 CTCF maintains retinoblastoma protein (pRb) and p53 gene promoter epigenetic status and tumour repression**

Further evidence for the role of CTCF in tumour suppression revolved around its interaction with the retinoblastoma protein (pRb). The protein product of the *Rb* gene negatively regulates cellular activities including differentiation and senescence via *E2F* target genes; it is a well-established tumour suppressor and most human tumours are associated with down regulation of this gene (Khidr and Chen, 2006). There are suggestions that CTCF might be involved in the regulation of the *Rb* gene via maintaining it in an epigenetically regulated state (La Rosa-Vela'zquez *et al.*, 2007). La Rosa-Vela'zquez and colleagues showed that CTCF bound to *Rb* gene promoters in

HeLa cells. They also found associated decreased *Rb* promoter activity on site-directed mutagenesis of the CTCF binding site suggesting that CTCF was an activator at the *Rb* gene promoter. Further still, they showed that methylation of the CTCF binding site at the *Rb* promoter abrogated CTCF binding confirming that the effect of CTCF at that site was via an epigenetic mechanism.

The role of CTCF in tumour suppression is also supported by its association with *p53* gene which encodes the p53 tumour suppressor protein and is central to the cellular metabolism of the body; indeed 50% of all tumours have a mutation of this gene while a significant proportion of the remainder have some dysregulation in the gene's signalling pathways (Zuckerman *et al.*, 2009). CTCF binding has been demonstrated at the promoter of the *p53* gene where that binding stopped the spread of repressive histone marks like H3K9me3, H3K27me3 and H4K20me3 (Soto-Reyes and Recillas-Targa, 2010). These latter authors also demonstrated that CTCF depletion was associated with the acquisition of repressive histone marks and consequent *p53* promoter gene silencing in glioma cell lines (Soto-Reyes and Recillas-Targa, 2010). In preserving *Rb*-mediated negative regulation and maintaining p53-activated status, CTCF in part, contributed to a tumour suppressive role.

### **1.10.3 CTCF maintains epigenetic balance at the cyclin-dependent kinase inhibitor 2A locus (CDKN2A)**

The cyclin-dependent kinase inhibitor 2A (CDKN2A) locus encodes two proteins, p14<sup>ARF</sup> and p16<sup>INK4a</sup>, which respectively stabilise *p53* and prevent the inactivation of *Rb* proteins (Ouelle *et al.*, 1995; Ozenne *et al.*, 2010). Both p14<sup>ARF</sup> and p16<sup>INK4a</sup> have important actions in cell cycle arrest and cell senescence and their inactivation is associated with cell transformation (Ouelle *et al.*, 1995; Ozenne *et al.*, 2010). CTCF has been shown to bind to the promoter of p14<sup>ARF</sup> but this binding was abrogated by methylation in a p14<sup>ARF</sup> – negative osteosarcoma cell line (Rodriguez *et al.*, 2010). Demethylation with 5'-aza-2'-deoxycytidine (AZA) successfully reversed the lack of CTCF binding and furthermore

CTCF depletion prevented the re-activation of p14<sup>ARF</sup> after AZA treatment suggesting that CTCF was indispensable in the transcriptional activation of the p14<sup>ARF</sup> promoter (Rodriguez *et al.*, 2010). With respect to p16<sup>INK4a</sup>, CTCF has been shown to possess barrier function ~2kb upstream to the transcriptional start site of the *p16* gene and is associated with the transcriptionally active *p16* gene but not the inactive gene (Witcher and Emerson, 2009). CTCF knockdown induced by shRNA caused the spread of repressive histone marks to the *p16* promoter with subsequent silencing of that gene (Witcher and Emerson, 2009). Furthermore, both DNA methylation and loss of CTCF PARylation were found to be factors associated with loss of CTCF binding at the boundary region and subsequent *p16* gene silencing confirming the impact of epigenetic mechanisms at this gene locus (Witcher and Emerson, 2009). Taken together therefore, the action of CTCF at the promoters of the CDKN2A locus served to further stabilise and keep them activated to prevent cellular deregulation.

### **1.11 CTCF and the clinical breast cancer phenotype**

CTCF has been linked CTCF to various forms of clinical breast cancer. With respect to high penetrance breast cancer genes, excessive methylation of the *BRCA1* promoter is seen in up to 20% of sporadic forms of breast cancer and is associated with a decrease in *BRCA1* gene expression and tumorigenesis (Matros *et al.*, 2005). CTCF is known to have binding sites in the region of the *BRCA1* promoter and is possibly involved in insulator functions in that area and in conjunction with specific protein 1 (Sp1) keeps that gene free of methylation in the normal breast (Butcher *et al.*, 2004). Deletion and mutations at the chromosomal locus of the *CTCF* gene, in conjunction with cytoplasmic location of CTCF (as opposed to nuclear) and absent CTCF barrier elements around the *BRCA1* gene could result in an epigenetic hit and tumour formation (Aulmann *et al.*, 2003; Rakha *et al.*, 2004). Further evidence in support of the impact of



CTCF on *BRCA* genes was presented by Butcher and Rodenhiser (2007), who interrogated the relationship between CTCF, a DNMT3b methyltransferase and *BRCA1* gene on the one hand and *BRCA1* promoter methylation on the other hand. They found that alterations in methylation patterns were critical events in *BRCA1* gene inactivation and sporadic breast tumours suggesting that epigenetic mechanisms could be the link between CTCF and *BRCA 1* dysregulation in breast cancer.

Contrary to this conclusion however is a recent effort that determined CTCF promoter methylation status and the associated CTCF mRNA expression in sporadic breast cancer patients (Wang and Zhang, 2014). The authors used methylation-specific PCR, bisulfite sequencing PCR, and quantitative real-time PCR and found that CTCF gene methylation was surprisingly lower in breast cancer tissue compared to normal breast tissue. Furthermore they discovered that CTCF mRNA expression was lower in breast cancer tissue compared to normal breast tissue suggesting that malignant transformation in this setting was probably not directly linked to methylation processes at the CTCF promoter itself. The data from this report is not straightforward as promoter hypermethylation is generally associated with gene silencing and therefore decreased gene expression. That pattern was not seen in this report where though CTCF promoter methylation was lower in cancer tissue, CTCF mRNA expression was not higher in those transformed cells compared to normal breast tissue. Generalisable conclusions obviously cannot be drawn from this study as the sample size was small (62 patients) and cancerous malformation is a multifactorial process. Moreover, the impact of CTCF protein on breast cancer may not be directly related to the extent of its mRNA and protein expression as multiple PTMs together with protein partners could alter the resultant phenotype. In comparison to the findings of Butcher and Rodenhiser (2007), it is possible that the mechanisms of CTCF-associated tumour malformation in conjunction with *BRCA1* gene defects may be different to those breast cancers without

*BRCA1* gene alteration. The conflicting reports point to the need for more research in this area.

In the case of familial breast cancer Zhou *et al.* (2004) investigated 153 cases with familial non-*BRCA1* / *BRCA2* breast cancer for germ line mutations in the *CTCF* gene. This case control study, using denaturing high-performance liquid chromatography followed by cycle sequencing found only two sequence variants in five cases but the sequences occurred at the same frequency between cases and controls. They suggested therefore that *CTCF* gene mutations did not appear be important sources of dysregulation associated with familial breast cancer. With such a small sample size, the study of Zhou *et al.* (2004) could only serve to point towards an area for further research. Taken together however, these studies (Zhou *et al.*, 2004; Butcher and Rodenhiser 2007; Wang and Zhang, 2014) suggested that whether breast cancer is familial or sporadic, the mechanism by which CTCF impacts on the breast cancer phenotype could partially revolve around epigenetic processes.

Loss of chromosomal material at Chromosome 16q is common in breast cancer and the centromeric part of this chromosome commonly deleted in breast cancer also harbours the gene for CTCF (Rakha *et al.*, 2005; Rakha *et al.*, 2006). Evidence from studies involving loss of heterozygosity (LOH) and comparative genomic hybridization (CGH) suggested that chromosomal loss at this locus occurred at about the same frequency in lobular cancer *in situ*, and in low and intermediate ductal carcinoma *in situ* (DCIS) but the frequency of this loss is generally thought to be lower in invasive ductal carcinoma (Rhakha *et al.*, 2006). Using breast tissue from normal and lobular carcinoma *in situ*, Green *et al.* (2009), applied real time polymerase chain reaction (RT-PCR) and immunohistochemistry and confirmed that CTCF expression was significantly reduced in the LCIS samples compared to normal breast tissue, lending support to the fact that loss of genetic material at this locus could be an early event in breast tumorigenesis starting from the *in situ* stages (Green *et al.*, 2009). It is not clear however how reliable the RT-

PCR results in this paper are as RT-PCR data have proved to be non-reproducible leading to the development of guidelines for publication of RT-PCR data (Bustin *et al.*, 2009). Regarding the more invasive forms of breast cancer, there appears to be conflicting data. Some studies using immunohistochemistry, DNA amplified probes, allelic studies and mutation screen of the *CTCF* gene found no loss of CTCF protein and no direct correlation between CTCF expression and tumour type in invasive breast cancer samples (Aulmann *et al.*, 2003; Rakha *et al.*, 2005). Another report however using western blotting and anti-CTCF monoclonal antibodies that detected a PARylated form of CTCF, found an association between low expression level of a CTCF isoform and worse breast cancer prognostic indices in invasive ductal carcinoma (Docquier *et al.*, 2009). This report while suggesting that CTCF could be a proliferation factor based on its behaviour in primary cultures also speculated that a possible evolutionary functional change could be responsible for the association between high CTCF expression and better breast cancer prognostic indices. It is difficult to reconcile CTCF expression levels and breast cancer outcome in this paper. Taken together, differences in methodology and the detection of particular forms of modified CTCF protein could be responsible for the disparate results noted in these studies that have investigated the general association between CTCF expression and breast cancer type.

### **1.12 CTCF and the estrogen receptor (ER) $\alpha$**

In a bid to further explain the role of CTCF in breast cancer pathogenesis, efforts have been geared towards defining the relationship between CTCF, estrogen and ER $\alpha$ . Most breast cancers are driven by estrogen and exogenous estrogen (E2) has been shown to downregulate CTCF mRNA expression in the ER positive MCF7 breast cancer cell line (Del Campo *et al.*, 2014). These authors confirmed basal CTCF mRNA expression in this cell line and noted that the downregulating effect of exogenous estrogen became

statistically significant at higher E2 concentrations. Using the LASAGNA-Search software they confirmed the presence of a consensus sequence for the estrogen response element (ERE) on the CTCF promoter suggesting that the CTCF promoter could be a target for ER $\alpha$ . These findings tend to support the earlier report of Ross-Ines *et al.* (2011) who analysed ER / CTCF binding data, generated by ChIP-Seq in MCF7 cells, overlapped with a previously published dataset that identified genes upregulated or downregulated on E2 stimulation. The authors found that estrogen-downregulated gene regions were more likely to be co-bound by CTCF and ER than by either alone. They also found that CTCF and ER binding events colocalised and that there was significant ERE and CTCF motif enrichment in those colocalised regions. While suggesting that CTCF could influence ER binding to chromatin they confirmed a previous report that suggested that CTCF partitioned the genome into blocks that may or may not contain ER $\alpha$  binding regions and ER-regulated genes (Chan and Song, 2008). Furthermore, the forkhead protein (FOXA1 / HNF3 $\alpha$ ) which modulates ER $\alpha$  – chromatin interactions and is an absolute requirement for ER $\alpha$  binding to ER promoters (even in the absence of E2 binding) is itself negatively regulated by CTCF (Hurtado *et al.*, 2011). More evidence linking CTCF to ER $\alpha$  involved epigenetic mechanisms. They include the activity of histone deacetylase (HDAC), a protein that is recruited by both CTCF and ER $\alpha$ , which via the epigenetic mechanism of histone deacetylation leads to repression of gene expression (Lutz *et al.*, 2000; Kawai *et al.*, 2003). Moreover, E2 stimulation not only downregulated the expression of CTCF, it also lead to CTCF recruitment to the CDKN1c promoter and a phenomenon that epigenetically induced gene silencing possibly via methylation (Rodriguez *et al.*, 2011). While there appears to be substantial evidence describing the effects of ER $\alpha$  activity on CTCF, there is very little information regarding how alterations in CTCF gene expression impacts ER $\alpha$  expression in breast cancer cells. In effect, the functional role of CTCF on ER biology and specifically ER gene expression is yet to be clearly explored. Further elucidation of

the activity of CTCF with respect to ER regulation could augment the manipulation ER $\alpha$  and E2 in breast cancer treatment.

### 1.13 Conclusion

Breast cancer has evolved over the years and in many patients is now considered a chronic disease. This has been made possible by the improvement in management strategies available for patients with the condition. This situation resulted from extensive research work that has led to the identification of breast cancer as a heterogeneous disease with distinct clinical types that possess different management strategies. Despite this improvement it is still associated with considerable suffering and death. This emphasised the need to continue with research effort in the field to widen the pool of knowledge and tools available for patient management. Recent work discovered that the 11 zinc finger protein, CTCF, could modify the clinical phenotype of breast cancer (Docquier *et al.*, 2009). This protein is evolutionally conserved from invertebrates to man and is ubiquitously expressed in the genome. It is involved in an array of cellular processes including but not limited to gene expression, genome organisation, imprinting and X-chromosome inactivation. The exact mechanisms that governed its activity are not clear but could involve protein partners. It is thought to have significant general tumour suppressor role and acts in concert with *p53*, *pRb* and *hTERT*. Furthermore, it possesses an impact on breast cancer phenotype but the relationship between *CTCF* gene status, mRNA and protein expression profile in relation to different phenotypic expression of human breast cancer is not known. Also, CTCF has a complicated relationship with the estrogen receptor which drives the majority of breast cancer cases. There is however no information regarding the direct effect of CTCF on ER expression. In effect, the mechanism(s) through which CTCF exerted its effect in breast tumorigenesis is / are not known. As epigenetic mechanisms alone do not seem to

explain how CTCF is involved in human breast cancer, there is therefore a pressing need for further research in the area of mechanisms especially the involvement of protein partners in the association between CTCF and breast cancer. This question was investigated in this thesis via two related projects.

#### **1.14 Aims and Objectives**

The main objective of the work presented in this thesis was therefore to study the mechanistic involvement of CTCF in human breast cancer via its protein partners in a panel of five breast cancer cell lines derived from different forms of human breast cancer and possessing different hormonal phenotype and invasive potential. Two related projects were undertaken. The first project, drawing on the possible involvement of CTCF with proliferation, investigated the association of CTCF with known proliferation factors in breast cancer namely, Ki67 and the proliferating cell nuclear antigen (PCNA). The second project, a continuation of the first, noting that majority of human breast cancers are estrogen receptor positive, investigated a possible regulatory relationship between CTCF and the estrogen receptor (ER) in an estrogen receptor - positive breast cancer cell line.

## Chapter 2

---

### MATERIALS AND METHODS

#### 2.1 MATERIALS

##### 2.1.1 Breast cancer cell (BCC) cell lines

The breast cancer cell lines used in this thesis were selected to represent the major classes of human breast cancer possessing different hormone / HER2 phenotype and invasive potential as described in section 1.2 and are shown in Table 2.1. The hormone receptor profile, source and storage of the cell lines are also shown in Table 2.1. The immortalized normal luminal cell line 226LDM was generated in-house at Essex University using viral constructs carrying the modified T antigen, Tag (U19dl89-97), and hTERT (O'Hare *et al.*, 2001; Docquier *et al.*, 2009) and was a kind gift from Prof Klenova, Essex University, Colchester, United Kingdom. Cells were used for experiments on achieving no more than 70% - 80% confluence on incubation.

##### 2.1.2 Culture media

###### 2.1.2.1 Culture medium for MCF7, T47D and BT474 cells

MCF7, T47D and BT474 were grown and propagated in Dulbecco's Modified Eagle's Medium (DMEM) / Ham's F12) (Life Technologies, UK). It was supplemented with 2.1 mM L-Glutamine (Lonza, Switzerland), 10% v/v fetal bovine serum (GIBCO, UK) and gentamicin (PAA, Austria) at a final concentration of 50 µg / ml and was stored at 4°C.

###### 2.1.2.2 Culture medium for SKBR3 and MDA MB 231

SKBR3 and MDA MB 231 cells were grown and propagated in Dulbecco's Modified Eagle's Medium (DMEM) with low glucose (Life Technologies, UK). The medium was supplemented with L- glutamine (2.1 mM final concentration) (Lonza, Switzerland), 10%

v/v fetal bovine serum (GIBCO, UK) and gentamicin (PAA, Austria) at a final concentration of 50 µg / ml. It was stored at 4°C.

#### **2.1.2.3 Culture medium for LDM226 breast cells**

This cell line was grown in DMEM / F-12 (Lonza, Switzerland) supplemented with 5 µg / mL insulin, 1 µg / mL hydrocortisone, 20 ng / mL epidermal growth factor, 20 ng / mL cholera toxin (all from Sigma), 10% v/v fetal bovine serum (GIBCO, UK), and 50 µg / mL gentamicin (PAA, Austria) (Docquier et al., 2009). It was stored at 4°C.

#### **2.1.2.4 Luria Bertani broth (LB) and Luria agar (LA)**

This broth was prepared with 0.1% NaCl (Fisher, UK), 1% Bactotryptone (Fisher, UK) and 0.5% yeast extract (Fisher, USA) in distilled water. To prepare Luria agar, 2% bactoagar (Fisher, USA) was added to LB. The broth was stored at room temperature while the cast Luria agar was stored at 4°C.

#### **2.1.3 Reagents and Buffers /gels / solutions**

All reagents and buffers together with their composition used in this thesis are detailed in appendix sections 1 and 2.

#### **2.1.4 Antibodies**

Primary and secondary antibodies together with the experiments in which they were used are listed in appendix section 3.

#### **2.1.5 Plasmids, siRNA and biologic agents**

CTCF and empty vector (EV) cytomegalovirus-driven plasmid expression vectors were a kind gift from Prof Klenova's laboratory, University of Essex, Colchester. They are listed in appendix section 3.4.

#### **2.1.6 QPCR primers**

Primers used in QPCR experiments were designed with the kind help of Prof Klenova's laboratory, University of Essex, Colchester. They were obtained from ThermoScientific (UK) and stored at -20°C. The technical datasheet is attached in the appendix section 4.



Cell line	Original Derivation	Hormone receptor profile	Source	Storage	Invasion in Matrigel	References
<b>MCF7</b>	Metastatic pleural fluid	ER+/PR + HER2 -	European cell bank	-80°C	Weak	Soule <i>et al.</i> , 1973; Brookes <i>et al.</i> , 1973
<b>T47D</b>	Metastatic pleural fluid	ER+/PR + HER2 -	European cell bank	-80°C	Weak	Savouret <i>et al.</i> , 1991; Lacroix and Leclercq, 2004
<b>BT474</b>	Breast invasive ductal carcinoma.	ER+/PR + HER2 +	Prof E Klenova, UniEssex	-80°C	Moderate	Lasfargues <i>et al.</i> , 1979; Lacroix and Leclercq, 2004
<b>SKBR3</b>	Metastatic pleural fluid	ER -/ PR - HER2 +	European cell bank	-80°C	Moderate	Cailleau <i>et al.</i> , 1978; Lacroix and Leclercq, 2004
<b>MDA MB 231</b>	Metastatic pleural fluid	ER -/PR - HER2 -	European cell bank	-80°C	High	Cailleau <i>et al.</i> , 1978; Lacroix and Leclercq, 2004
<b>LDM 226</b>	Generated in-house	ER- PR-	Prof E Klenova, UniEssex	-80°C	Unknown	Docquier <i>et al.</i> , 2009

**Table 2.1 Tabular annotation of breast cancer cell lines used in this thesis.** Also indicated are their original source, hormone receptor and HER2 profile, storage and invasive potential.

## **2.2 METHODS**

### **2.2.1. Cell culture procedures**

#### **2.2.1.1 Reviving cells from storage**

Frozen breast cancer cells were rapidly thawed, diluted in appropriate medium and spun (Eppendorf centrifuge 5810 R, Germany) at 453x *g* for 5 minutes. The pellet was resuspended in 15 ml of appropriate complete growth medium, transferred to a 75cm<sup>2</sup> flask (LabMart, USA) and placed in a 5% CO<sub>2</sub> incubator at 37°C.

#### **2.2.1.2 Cell passaging and cell count**

Cells were passaged when they reached 70% - 80% confluence by first discarding the growth medium and washing with 2ml of 2mM ethylene diamine tetra acetic acid (EDTA) (Acros Organics, Belgium) to chelate calcium ions. Adherent cells were then released with 1ml 0.05 v/v trypsin (PAA, Austria) – 0.02 v/v EDTA solution after 5 minutes incubation at 37°C. Cells were spun (Eppendorf centrifuge 5810 R, Germany) at 453x *g* in complete medium for 5 minutes and the pellet resuspended in 10ml complete medium. Cells were counted using a haemocytometer (Neubauer chamber – Mansfield™, Germany) and 1x10<sup>6</sup> cells seeded per 75cm<sup>2</sup> flask, diluted with 10-15ml complete medium and incubated in 5% CO<sub>2</sub> at 37°C. Cells used for experiments were passaged a maximum of three times, discarded and replaced by fresh cells from the frozen stored stock.

#### **2.2.1.3 Freezing down cells**

For storage, 10<sup>6</sup> to 10<sup>9</sup> cells were placed in a 1.5 ml cryovial with an equal volume of freezing solution consisting 10% dimethylsulphoxide (DMSO), 40% fetal bovine serum (FBS) and 50% complete medium. The cryovial was kept at -80°C in a cryovial freezing container which allowed the temperature to drop off 1°C per hour for 24 hours and then transferred to a liquid nitrogen drawer (CryoService, UK) for long term storage.

### **2.2.2 Trypan blue test for cell viability**

A cell suspension was prepared from frozen as described in section 2.2.1. 200 µl of this suspension was placed in an Eppendorf tube containing 500 µl of 0.4% trypan blue solution (Sigma, UK) and 300 µl of Hanks' Balanced salt solution (GIBCO, UK). The solution was mixed thoroughly and allowed to stand for 10 minutes at room temperature. 10 µl of the solution was introduced into a counting chamber and cells counted using a haemocytometer (Neubauer chamber – Marienfeld <sup>TM</sup>, Germany).

### **2.2.3 Breast cancer cell lysates**

#### **2.2.3.1 Cell lysate for Western blot analysis**

Adherent breast cancer cells were trypsinised as previously described, counted using a haemocytometer (Neubauer chamber – Marienfeld <sup>TM</sup>, Germany) and placed in 1.5 ml Eppendorf tubes. They were centrifuged at 453x *g* for 5 minutes, resuspended in medium and counted as previously described. 20 µl of 2x lysis / loading buffer was mixed with 1 x10<sup>5</sup> cells and the solution immediately vortexed (Whitimixer <sup>TM</sup>, Fisherbrand) to resuspend the cell pellet. Heating at 95°C with a heating block (TECHNE, USA) for 5 minutes was done to disrupt hydrogen and ionic bonds and further augment the activity of mercaptoethanol in the lysis buffer by linearizing the proteins. Lysates were used immediately or frozen at -20°C for use at a later date.

#### **2.2.3.2 Cell extract for immunoprecipitation**

Cancer cell lines cells were allowed to achieve 70% - 80% single layer confluence in a 75cm<sup>2</sup> flask and washed with ice cold PBS x 1 twice. Adherent cells were then scraped off with a cooled plastic cell scraper (Fisher, Mexico) in 2 ml ice cold PBS x 1. The suspension was placed into an Eppendorf tube and spun (Eppendorf 5415R, Germany) at 453x *g* for 5 minutes at 4°C. The supernatant was discarded and 0.4 ml of freshly prepared immunoprecipitation (IP) lysis buffer added to the pellet, which was vortexed

and placed on ice for 15 minutes. It was subsequently spun at 15700x *g* for 15 min at 4°C (Eppendorf 5415R, Germany). The supernatant was aspirated into another Eppendorf tube and kept on ice for immediate use while the pellet was discarded.

#### **2.2.4 Bovine serum albumin protein assay**

Aliquoted and frozen concentrations (200, 100, 50, 25, 12.5, 6.25, 3.175 mg/ml) of BSA (Sigma, USA) standard, prepared with the buffer used to make cell lysate samples to be tested, were allowed thawed at room temperature and 10µl of each pipetted at the bottom of a labelled 96 well microtitre plate (BioRad, USA) in duplicate. A further 10 µl of distilled water and an equal volume of the buffer used to prepare lysates were pipetted to the bottom of the 8<sup>th</sup> and 9<sup>th</sup> wells in duplicate to serve as negative controls. Furthermore, 10 µl each of a 1:1 and 1:2 dilution of the cell lysate sample to be tested was placed at the bottom of the 10<sup>th</sup> and 11<sup>th</sup> wells also in duplicate. One part of the dye reagent concentrate (Bio-Rad, USA) was diluted with four parts of deionised water (Bio-Rad, USA) and 200µl of the solution pipetted to each of the eleven wells in duplicate and mixed gently by pipetting up and down. The plate was allowed to stand at room temperature for 30 minutes. It was subsequently placed in a spectrophotometer (BioRad, USA) and absorbance measured at 595nm. A graph plotting absorbance versus BSA standard concentrations was created and a best fit line drawn to determine protein concentration.

#### **2.2.5 Indirect Immunofluorescence procedure**

Growing breast cancer cells at 70% - 80% confluence were plated on sterilised round coverslips (Thermo Scientific, USA) in a well of a 12-well plate (SPL, Korea) in the appropriate growth medium. This step served to attach cells to a stable support to enhance handling. They were incubated in 5% CO<sub>2</sub> at 37°C overnight. Following removal of growth medium, cells were fixed by placing 0.5ml of 4% paraformaldehyde (PFA) (Sigma, UK) diluted in PBS onto coverslips and left to stand for 15 minutes at

room temperature. They were washed three times in 100 mM glycine solution seven minutes each time. Permeabilisation was achieved by incubating with 1 ml of 0.25% v / v Triton X-100 solution (ACROS Organics, Belgium) for 20 minutes at room temperature with gentle rocking. This step served to increase the interaction of antibody with intracellular contents while maintaining the integrity of the cell membrane (Khanna *et al.*, 2006). In order to achieve epitope / antigen retrieval, coverslips were first anchored unto a slide with a sealant (HENKEL <sup>TM</sup>, Germany) and immersed in 100 ml of 10mM citrate buffer (10 mM citric acid pH 6.0) and heated in a microwave (DeLonghi <sup>TM</sup>, Italy) at 900 watts for 5 minutes. They were cooled in tap water for a few seconds and washed in PBS x 1 for 10 minutes. The heat served to unfold proteins making epitopes more accessible to antibodies while the buffer solution ensured that the unfolded proteins retained that conformation (Fowler *et al.*, 2011).

Secondary antibodies were anti-mouse or anti-rabbit and all derived from goat (refer to materials in appendix section 3.2). Coverslips were initially blocked in 2% goat serum (Vector, USA) in PBS / 0.05% Tween / 1% bovine serum albumin (BSA) solution (Sigma, UK) with gentle shaking for two hours at room temperature. Overnight incubation of coverslips was done with primary antibody in PBS / 0.05% Tween / 1% BSA solution. Negative control coverslips were incubated in PBS / 0.05% Tween / 1% BSA solution with no primary antibody. Incubated coverslips were each washed three times, seven minutes each time with PBS / 0.05% Tween / 1% BSA solution. In the dark, each coverslip underwent further two hour incubation with secondary antibody conjugated to a fluorochrome. Cells were counterstained with 4', 6-diamidino-2-phenylindole (DAPI) (Invitrogen, USA) at a final concentration of 5 µg / ml to identify the cell nuclei. Following three washes each lasting seven minutes, coverslips were mounted unto microscope slides in glycerol-based mounting medium (Vector, USA) for fluorescence. Mounted cells were inspected under fluorescence imaging with the Olympus IX71 microscope (Olympus, Japan) using the Q imaging digital camera and 8

bit monochrome setting. The blue filter was used to identify cell nuclei stained by DAPI (excitation and emission wavelength cut-off of 341nm and 452nm respectively); the green filter identified fluorescein-tagged regions with an excitation and emission wavelength cut-off of 437nm and 515nm respectively; while the red filter detected rhodamine conjugation with excitation and emission wavelength cut-off of 555nm and 627nm respectively. Image merging was performed with the ADOBE Photoshop CS2 software (Adobe, USA). Cell protein expression – by immunofluorescence - was calculated as the number of cell nuclei expressing the protein in a given high power field relative to the total number of cell nuclei in that field and expressed as a percentage.

## **2.2.6 Sodium Dodecyl Sulphate (SDS) – PolyAcrylamide Gel Electrophoresis (PAGE) and western blot analysis**

### **2.2.6.1 Gel preparation and electrophoresis**

Resolving buffer (6-8%) was poured between the glass plates of the electrophoresis equipment (ThermoFisher, USA) with 0.5 ml distilled water placed on top to remove air bubbles and allow a uniform straight gel to form. It was left to polymerise for 15 minutes. Stacking buffer was poured after inserting a 9-well comb and left to stand for another 15 min. Combs were removed and wells (covered with running buffer) washed and then loaded with lysates. Electrophoresis was performed at 125 volts - 40 milliAmperes (mA) - 5 watts for two hours via the electrophoretic power supply (Amersham Biosciences, Sweden) for one gel. After electrophoresis the gel was incubated in running buffer plus 1% methanol for fifteen minutes prior to semi-dry transfer.

### **2.2.6.2 Semi-dry transfer**

Proteins were transferred from gels to polyvinylidene fluoride (PVDF) membrane (Millipore, USA) via sandwich. The sandwich consisted of PVDF membrane (soaked for 10 seconds in 100% methanol and thoroughly washed in water) placed under the gel and both bordered on top and below by squares of Whatman paper soaked in transfer

buffer. Methanol in the transfer buffer served to keep the gel at the same size as the gel has a propensity to absorb water; it also served to remove SDS from the proteins in the gel and improved the ability of the proteins to bind to PVDF membrane. The transfer was run at 100mA - 35 volts for 2 hours for one gel and 200mA – 35 volts (2 hours) for two gels. After blotting, the membrane was washed in TRIS-buffered saline.

#### **2.2.6.3 Blocking, primary and secondary antibody incubations**

To minimise non-specific interaction between the PVDF membrane and incubating antibodies, the membrane was incubated with blocking buffer (3% non-fat fresh milk [Marvel] in PBS x 1) with gentle rocking for 2 hours. It was subsequently incubated with primary antibody dissolved in blocking buffer for two hours at room temperature or overnight at 4°C with gentle shaking. Washing was then performed with wash buffer (PBS / 0.05% Tween) three times for 10 minutes each time and further incubation with secondary antibody - horseradish peroxidase (HRP) labelled was done for two hours at room temperature with shaking. The membrane was again washed three times for 10 minutes each time with wash buffer (PBS / 0.05% Tween).

#### **2.2.6.4 Blot development**

Detection of the signal was performed by incubating the PVDF membrane in 800 µl of enhanced chemiluminescence (ECL) solution A and an equivalent amount of ECL solution B (Uptima, France) for three minutes and the excess drained off. It was placed inside a cellulose acetate plastic folder, inserted into a cassette (Kiran <sup>TM</sup>, India) and exposed to an X ray film (Kodak, Japan) for an appropriate length of time in a dark room. While still in the dark room, the film was extracted and immersed in developer solution (Sigma, USA) until bands were visible, rinsed off in cool water for 30 seconds and placed in a fixer solution (Sigma, USA) for a further one minute. It was rinsed off again with water and allowed to dry.

#### **2.2.6.5 Stripping blot membranes**

Antibodies in blotted membranes were removed by placing the membrane in 100 ml of warm (55°C) strip buffer composite solution (4g SDS, 1.4ml  $\beta$ -mercaptoethanol, 1.51g TRIS in 200ml distilled water) and the container sealed with Saran<sup>R</sup> cling film (UK). The container was dipped in a warm (55°C) water bath for 15 minutes to maintain the temperature of container contents and was rocked every 5 – 10 minutes. The membrane was subsequently washed for 10 minutes with PBS containing 0.1% Tween-20 (PBS-T pH 7.5) and the incubation and wash steps repeated one more time. Finally the membrane was subjected to the western blotting procedure starting with the blocking phase.

#### **2.2.6.6 Silver staining**

The SDS-PAGE gel was placed in 100ml of a solution containing 10% acetic acid / 10% methanol and incubated at 4°C with shaking overnight. Further incubations at room temperature were performed in 100ml solution of 5% acetic acid / 50% methanol for 30 minutes and finally a further 30 minutes in 100ml of 50% methanol solution only. The gel was then washed in ultrapure water three times, 10 minutes each time with shaking, at room temperature. After washing, it was incubated with 100ml of sensitizer solution for 30 minutes at room temperature with shaking. Further washing at room temperature with ultrapure water consisting four washes of 10 minutes each was performed. Subsequent incubation with 100ml of 0.1% silver nitrate (0.1%) solution for 30 minutes at room temperature was carried out. Following three quick washes, 10 seconds each with ultrapure water, the gel was immersed fully in 100ml of developer, changed every three to four minutes until bands appeared or the background became too high. The reaction was stopped with destain after spots had fully developed.



#### **2.2.6.7 Coomassie blue staining**

Gels post-electrophoresis and post-transfer were washed in distilled water for 5 minutes. They were incubated in colloidal Coomassie blue solution (Fisher, UK) at room temperature with shaking for 5 minutes and then washed in distilled water every 5 – 10 minutes for about 30 minutes. They were then left in fresh distilled water overnight with shaking to display protein bands clearly.

#### **2.2.7 Immunoprecipitation assay (IP)**

The contents of one vial of Protein A Sepharose beads (Sigma, USA) were mixed with 500µl of 20% v/v ethanol. A 50µl aliquot of mixed Protein A Sepharose beads was then washed with 1ml of PBS x 1 and quickly spun (Eppendorf centrifuge 5415R, Germany) for 20 sec at 2300x *g*. The supernatant was removed and the beads washed and spun a total of three times at 2300x *g* for 20 seconds. Washed beads were incubated with primary antibody (20µg antibody / 20µl beads) for three hours on a rotor shaker at 4°C. The IP cell lysates were then prepared as described previously (section 2.2.3.2). 50µl of the IP lysate was collected as input for subsequent western blot analysis. In order to eliminate the effect of protein-bead interaction, the remaining cell lysate was incubated with washed protein A Sepharose beads for 30min at 4°C on a rotary shaker (Stuart <sup>TM</sup>, UK). It was spun (Eppendorf centrifuge 5415R, Germany) at 100x *g* for 2 minutes at 4°C. The cleared cell lysate was collected and beads stored / discarded. A further 50µl of the cleared cell lysate was drawn and kept for western blot analysis to test the efficiency and specificity of the co-IP experiment. The cleared cell extract was placed in the antibody – beads mix and incubated overnight on a rotor shaker at 4°C. Post incubation, the IP reaction was spun at 100x *g* for 2 minutes at 4°C. The supernatant was removed and beads washed at 400x *g* for 1 minute at 4°C three times. The pelleted beads were lysed in 60µl of 2x SDS lysis / loading buffer. In order to check the efficiency

of the IP reaction, 10µl each of input; precleared cell lysate; IP reaction supernatant; and first, second and third washes of IP beads supernatant respectively was added to an equal volume of 2x SDS lysis buffer in separate microcentrifuge tubes. Together with the IP beads they were vortexed and boiled at 95°C for 5 minutes. They were subsequently loaded onto 6% - 8% gels for SDS-PAGE followed by western blotting.

## **2.2.8 RNA – based procedures**

### **2.2.8.1 RNA extraction from cells**

To lyse and release RNA from incubated cells, culture medium was discarded and 1ml of TRIsure (Bioline, UK) placed into each well of a 12 well plate and incubated at room temperature for 5 minutes with gentle rocking and pipetting to detach cells. In order to separate the three different phases into which lysed cells partition, samples were transferred to Eppendorf tubes and 0.2 ml of chloroform added. Tubes were shaken vigorously by hand for 15 seconds and then incubated at room temperature for 3 minutes. Centrifugation at 12000x *g* (Eppendorf centrifuge 5415R, Germany) for 15 minutes at 4°C was subsequently performed. The colourless upper aqueous phase (containing RNA) was transferred to another tube while avoiding the interphase. To precipitate RNA out of solution 0.5ml of ice cold isopropanol was added to the aqueous phase and this solution was incubated at room temperature for 10 minutes. It was then centrifuged at 12000x *g* (Eppendorf centrifuge 5415R, Germany) for 15 minutes at 4°C. The supernatant was removed and pellet washed in 1ml of 75% ethanol by inverting the tube multiple times. This served to remove remaining traces of guanidinium, an antiribonuclease. Further centrifugation at 7500x *g* (Eppendorf centrifuge 5415R, Germany) for 5 minutes at 4°C was done, supernatant was then drawn and discarded while the pellet was allowed to air dry for at least 10 minutes. The pellet were then dissolved in 45µl of RNase free water (Fisher, USA) and incubated for 10 minutes at 55°C. To remove contaminating DNA, 5µl of 10x Turbo DNase buffer (1x final) (Ambion,

USA) and 1  $\mu$ l of Turbo DNase (2U/ $\mu$ l) (Ambion, USA) were added, mixed and incubated at 37°C for 30 minutes. 0.1 (10%) volume DNase inactivation agent (Ambion, USA) was added and incubation done for 2 minutes at room temperature (RT). It was centrifuged at 10,000x *g* (Eppendorf centrifuge 5415R, Germany) for 2 minutes at 4°C. The supernatant was drawn and placed in a new tube and frozen at -20°C.

#### **2.2.8.2 RNA quality assessment using the AGILENT 6000 Bioanalyser**

The quality of isolated RNA was assessed with the RNA 6000 LabChip kit (Agilent <sup>TM</sup>, Germany). Before use, the ladder (which acts as the reference for data analysis) was denatured for 2 minutes at 70°C, aliquoted into 1  $\mu$ l samples and stored at -80°C. For regular use, the gel mix reagents were equilibrated to room temperature for 30 minutes then 550  $\mu$ l of RNA 600 Nano gel was placed in spin columns provided by the kit and centrifuged for 10 minutes at 1500x *g*. Aliquots of 65  $\mu$ l were prepared and stored at 4°C to be used within one month. 1  $\mu$ l of RNA 6000 Nano dye concentrate was then added to the 65  $\mu$ l aliquot of filtered gel, vortexed thoroughly and centrifuged for 10 minutes at 13000x *g* while protecting it from light. After cleaning the Bioanalyzer 2100 electrodes – according to manufacturer's instructions, the RNA Nano chip was placed on the chip priming station and 9  $\mu$ l of the gel-dye mix pipetted at the bottom of the well marked 'G'. Using the plunger on the priming station the gel was dispersed across the chip. A further 9  $\mu$ l of the gel-dye mix was added into two other marked wells. 5  $\mu$ l of the RNA 6000 Nano marker was subsequently pipetted in the well that is marked with a ladder symbol and into each of the 12 sample wells (numbered 1-12). 1  $\mu$ l of the RNA ladder was pipetted to the well, marked with the ladder symbol and 1  $\mu$ l of RNA samples was placed in each sample well. The chip was then vortexed (Vortexer, Agilent, Germany) for 60 seconds at 2400 rpm (vortex "set point") and immediately inserted into the Bioanalyzer 2100. The selected program to run the chip was "Eukaryotic RNA nano series II." A successful ladder run has seven sharp peaks namely, one marker and six RNA peaks in

the electropherogram. A successful total RNA run was characterized by one marker peak and two ribosomal peaks (18S and 28S).

## **2.2.9 Plasmid DNA procedures**

### **2.2.9.1 Bacterial cell transformation using DH5 $\alpha$ <sup>TM</sup> competent cells**

A tube of DH5 $\alpha$ <sup>TM</sup> *E. coli* bacterial cells was retrieved from -80°C and thawed on ice. For each transformation reaction, 50  $\mu$ l of DH5 $\alpha$ <sup>TM</sup> cells was placed in a 1.5 ml microcentrifuge tube and 1ng – 10ng (1  $\mu$ l – 5  $\mu$ l) of DNA added. The solutions were mixed and then placed on ice for 30 minutes. It was subsequently heated at 42°C for 20 seconds without shaking and then placed on ice for 2 minutes. 950  $\mu$ l of pre-warmed Luria broth was added to the tube and was incubated at 37°C for one hour with shaking at 225 rpm. Two different volumes of the transformation reaction were collected and spread on two separate prepared agar plates (with the appropriate antibiotic) and left to dry in a hood for 20 minutes. The agar plates – wrapped to prevent desiccation - were incubated overnight at 37°C.

### **2.2.9.2 Preparation of bacterial culture**

In order to make bacterial broth, 5ml Luria broth (and 5  $\mu$ l of appropriate antibiotic – 100  $\mu$ g /  $\mu$ l for both kanamycin and ampicillin) was placed in a 10ml Grainer<sup>R</sup> tube. A single colony of bacteria growing on the agar plate was picked with a sterile yellow tip and the tip placed in the Grainer tube. This tube was then incubated with shaking (225 rpm) (MaxQ, ThermoScientific, USA) overnight at 37°C. Bacterial growth was evidenced by a cloudy medium.

### **2.2.9.3 Plasmid DNA minipurification**

1.5ml of bacterial culture was placed in a centrifuge tube and spun at 16,100x *g* (Centrifuge 5415R, Germany) for 1 minute. The supernatant was discarded and dry

pellet resuspended in 200  $\mu$ l of ice cold Solution 1 (Tris-HCl, pH 8; 10mM EDTA; 100  $\mu$ g / ml RNase A) and vortexed vigorously. In order to release chromosomal DNA, 200  $\mu$ l of solution 2 (200 mM NaOH; 1% SDS) was added; the tube inverted four to six times to mix contents gently and placed on ice for 5 minutes. Chilled solution 3 (3M potassium acetate, pH 5.5) was then placed in the tube to precipitate chromosomal DNA, proteins and carbohydrates. The tube was centrifuged (Centrifuge 5415R, Germany) for 10 minutes at 16,100 x g and the supernatant - containing plasmid DNA - retrieved. To precipitate plasmid DNA, 600  $\mu$ l of isopropanol was added to the supernatant and incubated at room temperature - after vigorous mixing - for 20 minutes. The tube was then centrifuged for 10 minutes at 16,100x g. The supernatant was discarded and pellet resuspended in 200  $\mu$ l of sterile water with vigorous vortexing and the tube placed on ice. To further remove contaminants from the plasmid DNA, 200  $\mu$ l of phenol was added to the tube and vortexed vigorously. To achieve partitioning, the tube was centrifuged at 16,100x g for 5 minutes at room temperature (Centrifuge 5415R, Germany). The aqueous phase was removed into a new tube into which was placed 200  $\mu$ l of chloroform. This stage served to further partition impurities into the lower fraction. In order to remove cellular and histone proteins bound to DNA, the aqueous phase was transferred to a fresh tube and 50  $\mu$ l of 3M sodium acetate (pH 5.5) and 350  $\mu$ l of 100% ethanol added. The sodium acetate also helped preserve a somewhat alkaline pH, maintaining the solubility of DNA. A 10-minute centrifugation (room temperature) at 16,100x g was done and the supernatant removed and discarded. 400  $\mu$ l of 70% ethanol was added to the pellet to wash it, vortexed and centrifuged (Centrifuge 5415R, Germany) at 16,100x g for 5 minutes. The supernatant was discarded and the pellet air-dried for about 10 minutes. It was then resuspended in 30  $\mu$ l of sterile water and stored at -20°C.

#### **2.2.9.4 DNA quantitation with ultraviolet (UV) spectrophotometry (Nanodrop)**

In order to assess the quantity of DNA present in the material obtained from plasmid minipreparation, the Nanodrop spectrophotometer was first blanked with 1.5 µl of RNase free water. Subsequently 1.5 µl of each sample was placed on the cuvette and DNA concentration read off the spectrophotometer. The procedure ended with another blank reading to clean out the cuvette.

#### **2.2.9.5 Plasmid restriction enzyme digestion and agarose gel electrophoresis**

Agarose (1%) solution was prepared in TAE buffer. The bottle was covered loosely and heated in a microwave until agarose was well dissolved. It was cooled and while still liquid was poured into a gel-casting tray (ThermoFisher, USA) with combs already inserted and allowed to stand for 30 minutes. The tray was transferred to a horizontal electrophoresis tank (ThermoFisher, USA) and covered with TAE buffer solution. Samples (including loading buffer) already treated with digestion enzymes and SYBR green were loaded into the wells and electric field of 100V applied for one hour - in the dark. Nucleic acid bands post electrophoresis were visualised with the Odyssey infrared scanner (Licor, UK).

#### **2.2.9.6 Plasmid extraction using the Endofree plasmid Maxiprep kit (QIAGEN <sup>TM</sup>)**

After confirming the plasmids via minipreparation, enzyme digest and gel electrophoresis, large scale (maxi) preparation of the plasmids was performed with the endofree plasmid maxi kit (Qiagen <sup>TM</sup>). First, 250 ml of LB containing 250 µl of the appropriate selective antibiotic was prepared. It was then inoculated with 500 µl of bacterial broth obtained during the minipreparation stage and incubated at 37°C overnight for 12 hours with shaking at 300 rpm. Cells were then harvested by centrifugation at 6000x g for 15 minutes at 4°C and resuspended in 10 mls of buffer P1. A further 10 ml of buffer P2 was added to the tube and mixed thoroughly by inversion and the solution incubated for 5 minutes at room temperature. The incubated cell lysate

was further mixed with 10ml of chilled Buffer P3 and poured into the barrel of a prepared QIA filter Cartridge. The lysate was incubated in the barrel of the cartridge at room temperature for 10 minutes and then filtered into another tube by applying the plunger. Subsequently, 10% (2.0 ml) of Buffer ER was added to the filtered lysate, mixed by inversion and incubated at room temperature for 30 minutes. During this period, running 10ml of Buffer QBT through the column by gravity equilibrated the Qiagen-tip 500. The filtered lysate was then poured into the QIAGEN-tip and allowed to flow through the embedded resin by gravity. Subsequently the QIAGEN-tip was washed two times with Buffer QC to remove impurities and DNA trapped in the resin was eluted with Buffer QN. DNA was then precipitated out of the eluate by mixing with room-temperature isopropanol and centrifuging at 15000x *g* for 30 minutes at 4°C. The supernatant was decanted and DNA pellet washed in 70% ethanol (room-temperature) with further centrifugation at 15 000 x *g* for 10 minutes. The resultant pellet was air dried, redissolved in 500 µl of Buffer TE and stored at -80°C.

#### **2.2.10 Transfection assays**

##### **2.2.10.1 Transfection assays with plasmid expression vectors**

Incubated cells on achieving 40% – 80% single layer confluence were trypsinised and counted. An appropriate density of cells was diluted in antibiotic-rich medium and placed into each well of a 12 well plate. The cells were incubated with 5% CO<sub>2</sub> overnight at 37°C. Transfection was performed on achieving 40% to 70% cell confluence. To prepare the transfection complexes, plasmid vector and transfection reagent (Attractene, Qiagen) suspensions were prepared individually using antibiotic - free medium and left to stand for 5 minutes at RT. The plasmid solution was mixed with attractene transfection reagent and incubated for a further 20 minutes at room temperature. Complete medium was added to make a final volume of 1000 µl for each well of a 12 well plate. Culture medium from the incubated cells was removed and 1000 µl of prepared transfection complexes placed in each well. The cells were incubated with 5%

CO<sub>2</sub> at 37°C for a total of 48h and transfection complexes removed and replaced with fresh medium after 12 hours of incubation. At the end of incubation, transfected cells were either lysed for SDS-PAGE and subsequent western blotting or underwent RNA extraction for QPCR.

#### **2.2.10.2 Transfection assays with small interfering RNA (siRNA)**

Incubated cells on achieving 60% – 80% single layer confluence were trypsinised and counted. An appropriate density of cells was diluted in antibiotic-free medium and placed into each well of a 12 well plate. The cells were incubated with 5% CO<sub>2</sub> overnight at 37°C. Transfection was performed on achieving 40% to 70% cell confluence. To prepare the transfection complexes, siRNA (Dharmacon, ThermoScientific) and transfection reagent (DharmaFECT 1, ThermoScientific) suspensions were prepared individually using serum - and antibiotic - free media and left to stand for 5 minutes at room temperature. The medium was serum-free since complexing of siRNA and transfection reagent in serum lowers the efficiency of that process. The siRNA dilution was then mixed with the transfection reagent and incubated for a further 20 minutes at room temperature. Antibiotic-free medium (to limit cell cytotoxicity) was added to make a final volume of 1000µl for each well of a 12 well plate. Culture medium from the incubated cells was removed and 1000µl of prepared transfection complexes placed in each well. The cells were incubated with 5% CO<sub>2</sub> at 37°C for a total of 48h – 72h and transfection complexes removed and replaced with fresh medium (antibiotic-free for siRNA) after 12 hours of incubation. At the end of incubation, transfected cells were either lysed for SDS-PAGE and subsequent western blotting or underwent RNA extraction for QPCR.



## **2.2.11 Reverse transcription – polymerase chain reaction for (RT-PCR; QPCR) procedures**

### **2.2.11.1 Complimentary DNA (cDNA) synthesis**

1µg of RNA was denatured by heating at 70°C for 5 minutes and placed immediately on ice. It was mixed with 4 µl of 5 x cDNA synthesis buffers, 2 µl of 500 uM dNTP mix, 1 µl of anchored oligodT primers (500ng / µl), 1 µl of 0.5 µM Verso enzyme mix and 1 µl of reverse transcriptase enhancer, all from the Verso cDNA kit (ThermoScientific, UK). The reaction mixture was made up to 20 µl with RNase free water and incubated in a PCR machine (Genestorm, England) at settings consistent with the manufacturer's instructions. The cDNA sample was saved at -20°C for subsequent applications.

### **2.2.11.2 Preparing primers for quantitative PCR (QPCR)**

To prepare 100 µM solution of primer pair, the vial containing the dry primer powder was spun down and the recommended amount of RNase free water (from technical sheet – see appendix) was placed in the tube and mixed by inverting several times. The tube was kept on ice for 20 minutes, centrifuged again and stored at -20°C.

### **2.2.11.3 Standard curve determination for QPCR efficiency**

A serial log dilution of cDNA for QPCR was obtained from  $1 \times 10^2$  through to  $1 \times 10^6$  concentrations. QPCR was then performed as with the procedure in section 2.2.7.4 below. A standard curve was automatically generated by the thermal cycler (CFX Connect <sup>TM</sup>, Bio-Rad, USA).

### **2.2.11.4 Quantitative polymerase chain reaction (QPCR)**

To obtain enough samples for two wells (duplicate) of a 96 well plate, 5 µl of Kapa mastermix was mixed with 3 µl of diluted (1:5) cDNA, 3 µl of RNase – free water and 1µl of 10 µM concentration of the relevant primer. 5 µl of this solution was pipetted to the

bottom of each of two wells of a 96 well plate. The plate was labelled accordingly, centrifuged for one minute and placed in the thermal cycler (CFX Connect <sup>TM</sup>, Biorad, USA). The settings for the thermal cycler corresponded to the manufacturer's instructions for optimum temperatures for the Kapa mastermix (see appendix section 5).

#### **2.2.12 Liquid chromatography – mass spectrometry (LC – MS / MS)**

In order to perform LC – MS / MS, breast cancer lysates already subjected to immunopuification via immunoprecipitation, underwent SDS PAGE together with an IgG negative control. The gel was then stained with Commassie blue and was subjected to in-gel digestion. Gel digestion, preparation of samples and the actual mass spectrometry process was kindly carried out at the regional proteomics centre in the department of Biological Sciences, Essex University by Dr Metodi Metodiev and colleagues and as described in Alldridge *et al.* (2008).

## CHAPTER 3

---

### **RESULTS - Investigating CTCF protein partners in a panel of five breast cancer cell lines**

#### **3.1 Background**

##### **3.1.1 Cell proliferation and breast cancer**

The rate of proliferation of a cancer cell could determine its invasive and metastatic potential and in turn could be directly linked to its prognosis (Stuart-Harris *et al.*, 2008). A major part of invasive breast cancer research has focused on proteins that could define the rate of proliferation of a tumour and therefore suggest its malignancy potential (Van Diest *et al.*, 2004). Determining the aggressiveness of a tumour helps in the stratification of patients with regards to treatment options. Some of the techniques / factors that might have a role as markers of cell proliferation linked to prognostic outcome have been reviewed in Fitzgibbons *et al.* (2000) and Patani *et al.* (2013) and were summarised in table 1.3.

##### **3.1.2 CTCF and proliferation**

CTCF is a protein involved in an array of cellular activities including regulation of cell proliferation (section 1.6.3.2). An early study investigating the effect of CTCF on cellular proliferation in HEK 293 cells, a human embryonic kidney cell line, found that overexpression of CTCF markedly inhibited cell growth (Rasko *et al.*, 2001). These authors also found that cells of the 293 cell line which lacked functional *p53* and retinoblastoma (*Rb*) genes, when expressing CTCF, remained viable and either divided slowly or not at all, for up to a week. Since functional *p53* and retinoblastoma (*Rb*) genes are prerequisites for growth, they concluded that CTCF was an important part of the complex network of genes critical for proliferation and growth in these cells.

Furthermore, evidence for the involvement of CTCF in proliferation was presented in both normal cells and in breast cancer cells / tissue. With respect to normal cells, Heath *et al.* (2008) worked on T cells directly isolated from healthy mice thymus and showed that absent CTCF expression was associated with decreased  $\alpha\beta$  T cell differentiation and a block in the cell cycle that led to the production of small T cells. Associated with this proliferation block was an increased expression of *p21* and *p27*, which are major cell cycle inhibitors. Further evidence for the involvement of CTCF in proliferation in normal cells related to the findings of Li and Lu (2005) who investigated the functional role of *PAX6*, a transcription factor important in ocular development, in epithelial growth factor (EGF) - induced corneal cell proliferation. They discovered that increased CTCF expression, induced by EGF, led to an augmentation of corneal cell proliferation. The rise in CTCF expression was thought to be due to a rise in *Erk* signalling and the observed down regulation of *pax6* activity was linked to the action of CTCF at the *pax6* P0 promoter. Conversely, the authors observed that abolition of CTCF mRNA expression via siRNA knockdown resulted in the upregulation of *pax6* expression with subsequent decrease in corneal epithelial cell proliferation. This latter activity was observed irrespective of changes in EGF signalling suggesting that CTCF involvement in corneal cell proliferation was essential and probably not dependent on EGF. Taken together these reports showed that CTCF served to regulate cell proliferation in normal and immortalised cells though the exact mechanisms are still not clear.

In breast cancer cells, CTCF protein expression was shown to be partly regulated by poly (ADP-ribosyl)ation (PARylation) where poly ADP-ribose (PAR) groups imparted a negative charge on the CTCF protein, a change that altered its interaction with DNA (Farrar *et al.*, 2010). Another publication, linked loss of CTCF PARylation with breast cancer phenotype and cell proliferation (Docquier *et al.*, 2009). This latter paper, previously mentioned in section 1.11, interrogated the differential expression of two isoforms of CTCF protein – 180kDa and 130 kDa - in breast tumour tissue, a breast

cancer cell line and a normal breast epithelial cell line generated in-house (LDM 226). The authors discovered that both CTCF isoforms, CTCF-180 (poly ADP-ribosylated) and CTCF-130 (hypo ADP-ribosylated) were present in normal breast tissue but only the CTCF-130 isoform was expressed in breast tumours. They also found that when cells from normal breast tissue were cultured and divided *in vitro* the expression of CTCF transited from the CTCF-180 isoform to the CTCF-130. Conversely, the CTCF-180 isoform appeared following growth arrest in breast cell lines. On account of this characteristic, the authors suggested that the CTCF-130 isoform could be a proliferation marker. Furthermore, they showed that 87% of breast tumour tissue samples studied expressed the CTCF-130 variety and a reduction in the absolute expression level of this isoform was associated with worse breast cancer prognostic indices. Since the main difference between normal and cancer tissue is dysregulation of cellular proliferation, the authors concluded that the transition from the poly-ribosylated to hypo-ribosylated forms of CTCF could mark the onset of dysregulation of cellular proliferation and the breast cancer phenotype. Proliferation markers including Ki67, as shown in table 1.3, could have an evolving role in breast cancer prognostication and a suggestion that CTCF and the proliferation marker, Ki67, may colocalise in breast cancer tissue warranted further investigation as the Ki67 antigen is a proliferation marker that has prognostic significance in breast cancer (Stuart-Harris *et al.*, 2008).

### **3.1.3 Ki67 protein and proliferation**

The Ki67 protein was initially discovered as two isoforms (320kDa and 359kDa) derived from alternative splicing (Schlüter, 1993). Any difference in action between the two isoforms is not known. The fraction of Ki67 positive cells (Ki67 labelling index) in a tumour sample can indicate the degree of proliferation of the tumour and the greater the number of Ki67 positive cells the worse the clinical prognosis (Urruticoechea *et al.*, 2005). Ki67 protein expression has been shown by immunocytochemistry to be present only in active phases of the cell cycle (G1, S, G2, M) but not in the resting (G0) phase

(Scholzen and Gerdes, 2000; Yerushalmi *et al.*, 2010). The absence of Ki67 in resting / quiescent cells has been challenged by Bullwinkel *et al.* (2006) who detected low levels of Ki67 at sites linked to ribosomal RNA (rRNA) synthesis. Ki67 expression has been shown to be required for progression through cell division as evidenced by the halting of proliferation when its expression is inhibited by antisense nucleotides (Schlüter, 1993). The exact function of Ki67 is still unknown but there are suggestions that it could be involved in organising DNA; have architectural or structural roles in the nucleolus; or be involved in ribosomal RNA synthesis (MacCallum and Hall, 2000; Bullwinkel *et al.*, 2006; Rahmanzadeh *et al.*, 2007).

The prognostic role of this protein was studied in a meta-analysis based on 43 out of 85 studies and covering 15790 out of a total of 32,835 patients with breast cancer (Stuart-Harris *et al.*, 2008). The study revealed a strong correlation between high Ki67 expression levels and worse patient survival. Concerning the possible predictive role of Ki67 labelling index, two trials (International Breast Cancer Study Group [IBCSG] VIII and IX), assessed the impact of this index on response to adjuvant chemoendocrine therapy in patients with endocrine responsive tumours. While confirming that high Ki67 expression levels was associated with poor prognostic features, the trials did not however reveal a predictive role for the Ki67 index for chemoendocrine therapy relative to endocrine therapy alone in node negative patients (Viale *et al.*, 2008). An international clinical trial (POETIC – Trial of Perioperative Endocrine Therapy – Individualising Care) closed patient recruitment at the end of October 2013 and will among other issues assess whether Ki67 could be a predictor for relapse free survival (RFS) in individual breast cancer patients (Smith *et al.*, 2011). The results of this trial are yet to be published. In the neo-adjuvant setting, Yerushalmi *et al.* (2010), retrieved and analysed data from 12 neoadjuvant breast cancer trials involving chemotherapy and on account of evidence from this meta-analysis concluded that high expression levels of Ki67 was generally found to be a predictor for good response to chemotherapy regimens. This is

possibly because more rapidly dividing cells (which generally express high levels of Ki67) are more sensitive to chemotherapy (Wang *et al.*, 2011). Other proliferation factors like the proliferating cell nuclear antigen (PCNA) have also been evaluated and could have some impact on breast cancer prognostication (Stuart-Harris *et al.*, 2008).

#### **3.1.4 Proliferating cell nuclear antigen (PCNA), proliferation and cancer**

PCNA is the molecular coordinator in the core DNA synthesis machinery and its association with cell proliferation is primarily genetic (Majka and Burgers, 2004). The main genetic mechanism of PCNA involvement with proliferation relates to its function as a DNA sliding clamp which via attachment to replication factor C, tethers polymerases to DNA (Majka and Burgess, 2004). This attachment increases the speed and efficiency of DNA polymerases and the processing and joining of the Okazaki fragment during DNA synthesis (Moldovan *et al.*, 2007). PCNA activity in turn is regulated by posttranslational modifications including ubiquitination (Fox *et al.*, 2011). PCNA monoubiquitination is associated with the ability to bypass a DNA lesion while polyubiquitination prepares PCNA for degradation (Fox *et al.*, 2011). There is an association between PCNA and breast cancer prognosis as data sourced from eleven publications involving 2677 patients found that breast cancer tumours overexpressing PCNA were invariably associated with shorter overall survival (OS) and disease free survival (DFS) (Stuart-Harris *et al.*, 2008). Concerning its interactions, PCNA has multiple protein partners (Moldovan *et al.*, 2007) including PARP1, a known partner of CTCF (Frouin *et al.*, 2003; Farrar *et al.*, 2010). Moreover, unpublished data suggested that the two proteins (CTCF and PCNA) could colocalise in the MCF7 breast cancer cell line warranting confirmation. It is therefore possible that both CTCF and PCNA could directly interact in breast tumorigenesis.

### **3.2 Knowledge gap and hypothesis**

The exact mechanism(s) through which CTCF is involved in breast cancer is / are not known. To determine the mechanism of action of CTCF in breast cancer a study of possible protein interacting partners was undertaken. The activity of a protein could be derived from its interaction with another protein whose functions are better known (Bolte and Cordelieres, 2006). In deriving a hypothesis, it was considered that Ki67 and PCNA are known proliferation factors. CTCF could also be a proliferation factor (Docquier *et al.*, 2009). Next, CTCF and PCNA are known partners of PARP1. Lastly, CTCF might colocalise with PCNA in a breast cancer cell line. This chapter therefore hypothesized that CTCF may directly interact with Ki67 and PCNA in breast cancer cells.

### **3.3 Objectives of this chapter**

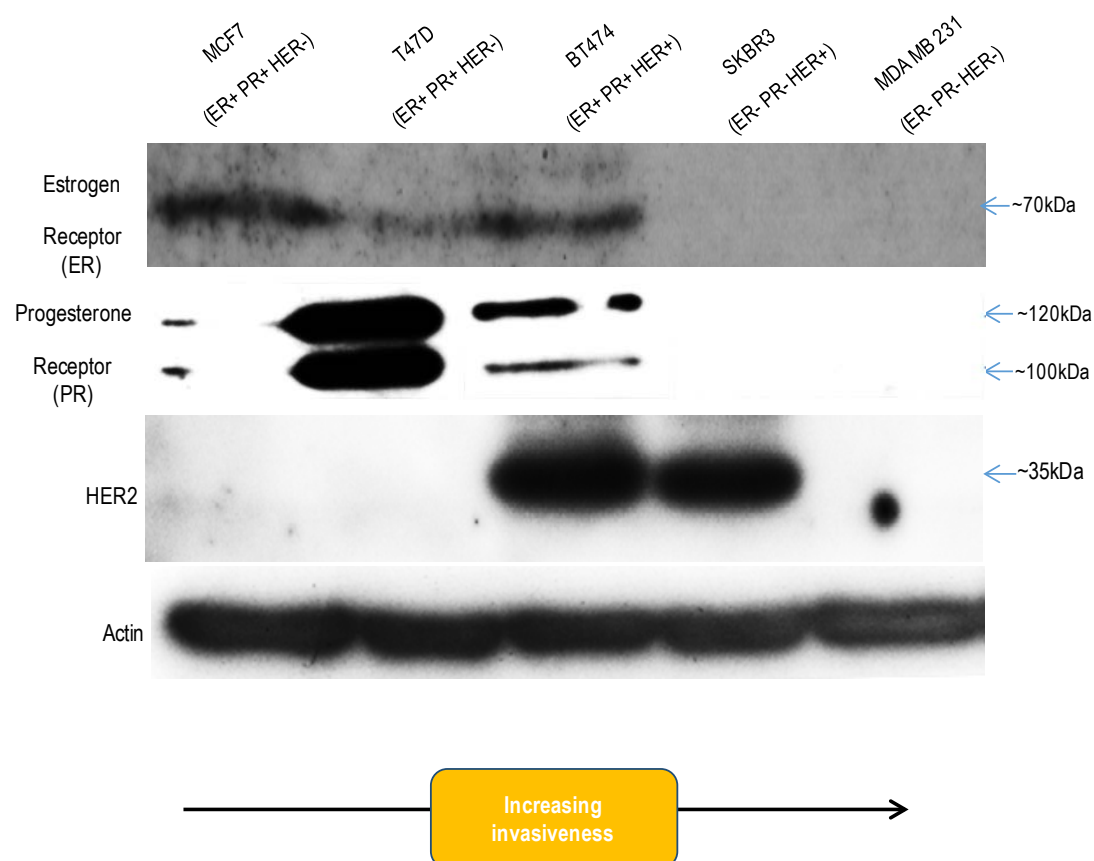
The main objective of this part of this chapter was to elucidate further the mechanistic role of CTCF in breast cancer via a possible interaction with two well-known breast cancer proliferation markers, Ki67 and PCNA. The first objective involved co-localisation studies of CTCF with Ki67 and PCNA performed via indirect immunofluorescence in a panel of five breast cancer cell lines possessing different hormone receptor and invasive phenotypes. Following co-localisation, co-immunoprecipitation assays and mass spectrometry were performed to determine whether CTCF was physically bound to either Ki67 or PCNA in the different breast cancer cell lines studied.



### **3.4 Results: Novel CTCF protein partners in five breast cancer cell lines**

#### **3.4.1 Confirmation of estrogen receptor (ER), progesterone receptor (PR) and HER2 receptor expression status in breast cancer cell lines**

There are numerous breast cancer cell lines now available from the American and European cell banks for use in research. In order to better understand the action of CTCF in relation to breast cancer, five different breast cancer cell lines were selected to reflect differences in the hormone receptor expression profile and invasive potential of the disease. To determine whether the hormone receptor (ER and PR) and HER2 expression profile of the cell lines used in this thesis matched the expected in the American Type Culture Collection (ATCC) cell bank, western blot analysis of total protein expression of hormone receptors (ER, PR) and HER2 in the five breast cancer cell lines was performed. Lysates (50 µg) of MCF7, T47D, BT474, SKBR3 and MDA MB 231 breast cancer cell lines were generated and resolved by SDS PAGE. The blotted membranes were probed sequentially with antibodies to estrogen receptor (ER)  $\alpha$ , progesterone receptor (PR), human epidermal receptor (HER) 2 and actin, the latter to determine protein loading. The results in figure 3.1 show MCF7 cells with moderate ER and relatively weak PR expression while being HER2 negative. ER expression in T47D cells was weak with very strong PR levels and negative HER2 expression. BT474 cells showed weak ER expression, moderate PR and strong HER2 levels. The same strong HER2 expression is observed with SKBR3 cells that also demonstrate negative ER and PR expression. MDA MB 231 is negative for both hormone receptors and HER2. Actin loading was shown to be uniform confirming equal protein loading across all cell lines. All cell lines revealed expected hormone receptor and HER2 expression (American Tissue Type Collection).

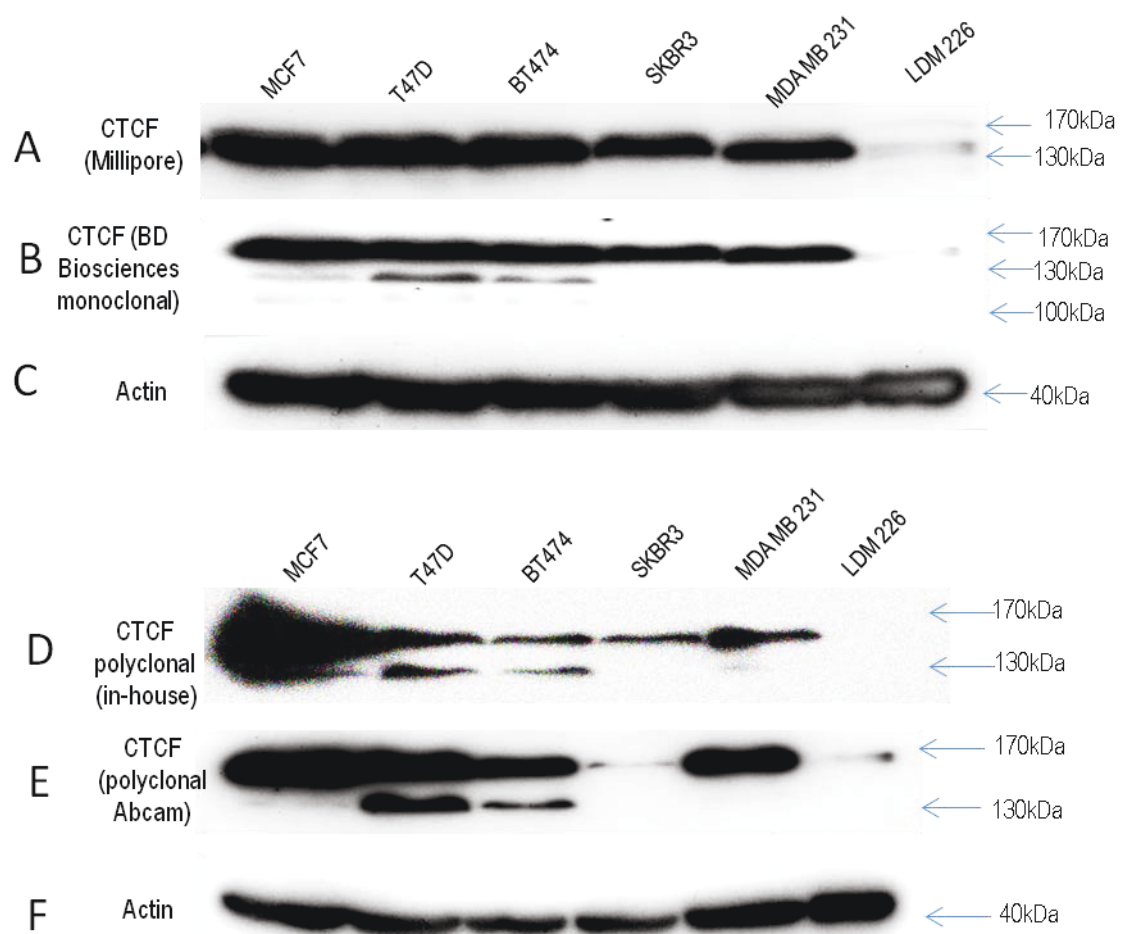


**Figure 3.1: Hormone receptor (ER and PR) and HER2 expression profile of five breast cancer cell lines.** Cell lysates (50µg) were prepared from MCF7, T47D, BT474, SKBR3 and MDA MB 231 cells. Protein expression levels of ER, PR, HER2 and actin were assessed by SDS-PAGE and western blotting. The concentrations of primary and secondary antibodies are stated in appendix section 7.2. The Figure shows ER and PR expressed as expected in the MCF7, T47D and BT474 cell lines while they were negative in SKBR3 and MDA MB 231 cell lines. HER2 expression was identified in BT474 and SKBR3 also as expected. Actin loading was uniform across all cell lines.

### 3.4.2 Total CTCF protein expression in different breast cancer cell lines

To date there are no comprehensive comparative studies looking at the expression of CTCF in breast cancer cell lines that possess different hormone receptor / HER2 and invasive phenotypes. Cell lysates (50µg) of MCF7, T47D, BT474, SKBR3 and MDA MB 231 breast cancer cell lines and of a normal epithelial breast cell line, LDM 226 (a kind gift from Prof Klenova, Essex University), were therefore generated to assess total CTCF protein expression. The lysates were resolved by SDS PAGE and blotted membranes probed with three commercially available antibodies to CTCF and one anti-CTCF [N terminal] antibody (a kind gift, generated from Prof Klenova's laboratory, Essex University). These antibodies to CTCF were chosen as CTCF is thought to exist in two main isoforms (180kDa and 130kDa) (Docquier *et al.*, 2009) and data not published (personal communication with F. Docquier) suggested that the anti CTCF (N terminal) antibody might distinguish between the two isoforms. The results shown in figure 3.2A and 3.2B revealed near uniform CTCF protein expression (somewhat lower in SKBR3 and MDA MB 231 cell lines), detected with the monoclonal anti CTCF antibodies (Millipore and BD Biosciences), at the manufacturers' published molecular weight of 140kDa. The antibody from BD Biosciences also revealed faint protein bands between 100kDa and 130kDa in the ER positive cell lines that could be artefactual. The polyclonal antibodies (N-terminal and Abcam, figures 3.2D and 3.2E) identified two CTCF isoforms. These were the ~160kDa present in all cell lines though expressed less in BT474 and SKBR3 cells and the ~130kDa which has variable expression in the hormone receptor positive cell lines but no expression in the hormone receptor negative and LDM 226 cell lines. Taken together, the disparate isoforms identified pointed towards the ability of polyclonal anti CTCF antibodies to detect more CTCF isoforms and confirmed the findings of Zhang *et al.* (2004) who suggested that there were multiple CTCF isoforms in HeLa cells whose expression was dependent on the phase of the cell cycle. The results in figure 3.2 in addition suggested that CTCF isoform expression

might also be cell type-dependent seeing that the lower molecular weight isoform detected by the polyclonal antibodies were not present in SKBR3, MDA MB 231 and LDM226 cells. Interestingly, there was no protein expression band at or above 170kDa in all the cell lines. There was also almost no CTCF expression in the normal epithelial cell line (LDM226) suggesting that CTCF expression is more preponderant in breast cancer cells compared to normal breast cells. Furthermore, the finding of a strongly expressed higher molecular weight CTCF isoform (~160kDa) in MCF7 cells is at variance with report of Docquier *et al.* (2009) who suggested that MCF7 cells predominantly expressed the 130kDa isoform. On account of its ability to detect CTCF protein (~140kDa) across the cell lines almost uniformly, the anti-CTCF antibody from BD Biosciences was used in all aspects of western blotting in this thesis.

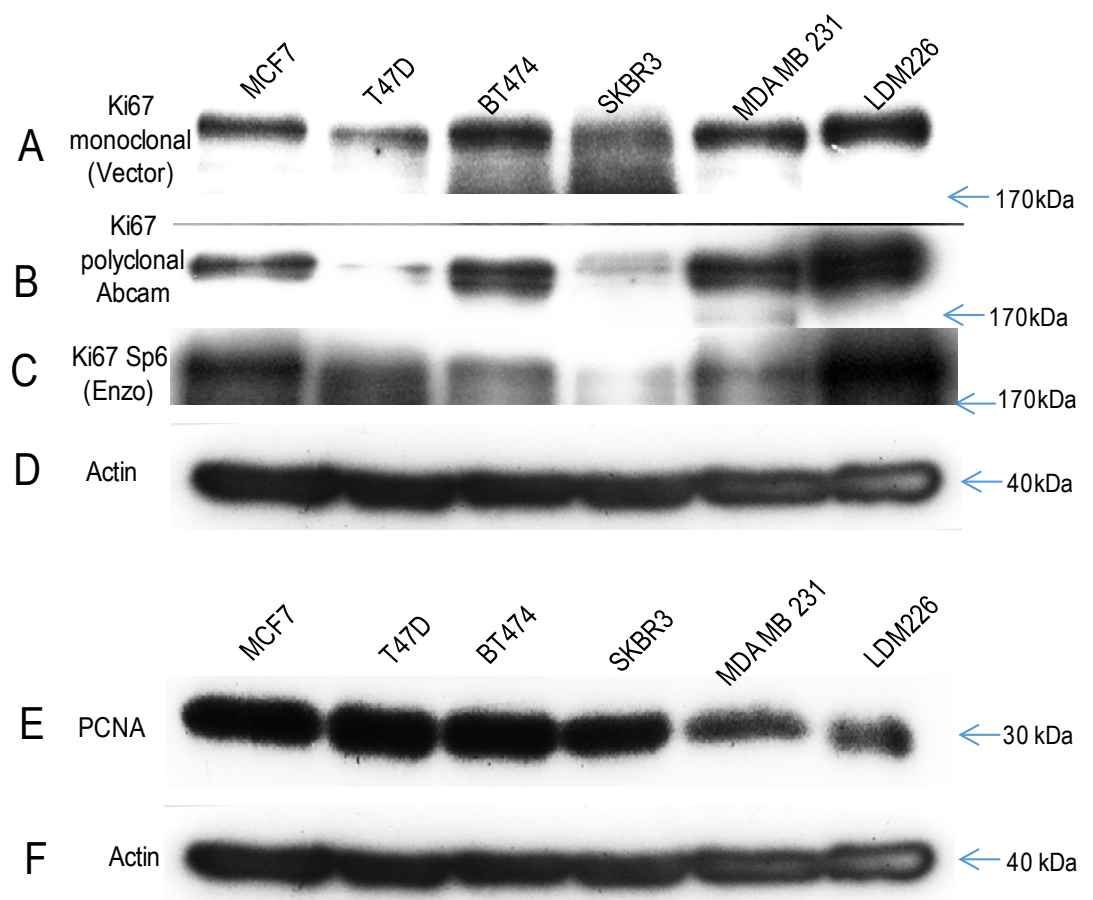


**Figure 3.2: Total CTCF protein expression in a panel of five breast cancer cell lines and one normal breast epithelial cell line.** 50µg cell lysate of MCF7, T47D, BT474, SKBR3, MDA MB 231 breast cancer cells and the normal LDM 226 breast epithelial cell line cells were loaded onto an 8% acrylamide gel, run at standard settings, resolved by SDS PAGE, blotted and probed sequentially. Shown are uniform CTCF expression bands detected with monoclonal anti-CTCF antibodies (though lower in SKBR3 and MDA MB 231 cell lines); low CTCF expression in BT474, SKBR3 and MDA MB 231 detected with the polyclonal anti-CTCF (N terminal) and almost absent CTCF expression in SKBR3 on probing with the polyclonal anti-CTCF antibody from Abcam. There was essentially no CTCF detected in the LDM226 normal breast cell line. Actin loading was relatively uniform in all cell lines.

### **3.4.3 Differential expression of proliferation markers, Ki67 and PCNA, in a panel of different breast cancer cell lines**

The proliferation status of a cell can be assessed by the expression level of proliferation markers such as Ki67 (Urruticoechea *et al.*, 2005). Ki67 protein expression is often assessed by immunohistochemistry (IHC) but there is limited data regarding its total expression using western blotting (WB) and certainly no information on its differential expression in breast cancer cell lines via IHC or western blotting. PCNA is important in DNA transcription and cellular proliferation where it functions as a DNA clamp (Moldovan *et al.*, 2007). While it is clear that different breast cancer types could have different proliferation rates, it is not known whether PCNA expression varies among these cell lines and if that expression has a temporal relationship to CTCF and / or Ki67 expression. As with Ki67, there is no information on the differential expression of PCNA in breast cancer cell lines via western blotting. Experiments were therefore performed to determine Ki67 and PCNA expression on western blotting in a panel of five breast cancer cell lines and a normal breast epithelial cell line. Since the anti-Ki67 antibodies available are mostly optimised for IHC, three different anti Ki67 antibodies were used to determine Ki67 expression via western blotting in this thesis. Of the three, only the Sp6 antibody epitope is characterised and is located within the C terminus of the Ki67 protein (anti-Ki67 antibody [Sp6]). One commercial antibody to PCNA, which has been well optimised for PCNA detection, was used to determine PCNA expression. The results of Ki67 protein expression revealed relatively similar protein expression across the breast cancer cell lines using the antibody from Vector (figure 3.3A). The Abcam anti Ki67 antibody detected moderate expression levels in MCF7 and BT474 cells (figure 3.3B). This expression is stronger in MDA MB 231 cells and even more so in the LDM 226 cell line suggesting a higher cell proliferation rate. The expression in T47D and SKBR3 cell lines is minimal compared to the other cell lines suggesting that the epitope to this antibody may be masked. Also shown in figure 3.3B are two protein bands running

closely together which might represent the better-known 342kDa and 395kDa isoforms of Ki67 protein. The Sp6 antibody detected a similar pattern of Ki67 expression to the Abcam antibody where low expression levels were observed in the T47D and SKBR3 cell lines (figure 3.3C). The LDM 226 cell line, immortalised from normal breast epithelial cells, is expected to have a relatively lower proliferation rate compared with the breast cancer cell lines and therefore lower Ki67 expression (Gerdes *et al.*, 1991). It was surprising to observe that Ki67 expression level in this cell line was greater than in the cancer cell lines. For PCNA, the results in figure 3.3E showed that it is strongly expressed in all cell lines but to a lesser level in MDA MB 231 and LDM 226 cells. Looking at figure 3.3 as a whole, there appeared to be a generally higher Ki67 expression together with a lower PCNA expression in the MDA MB 231 and LDM 226 cell lines. It is not clear whether there is a mutual association between Ki67 and PCNA proteins in these two cell lines.



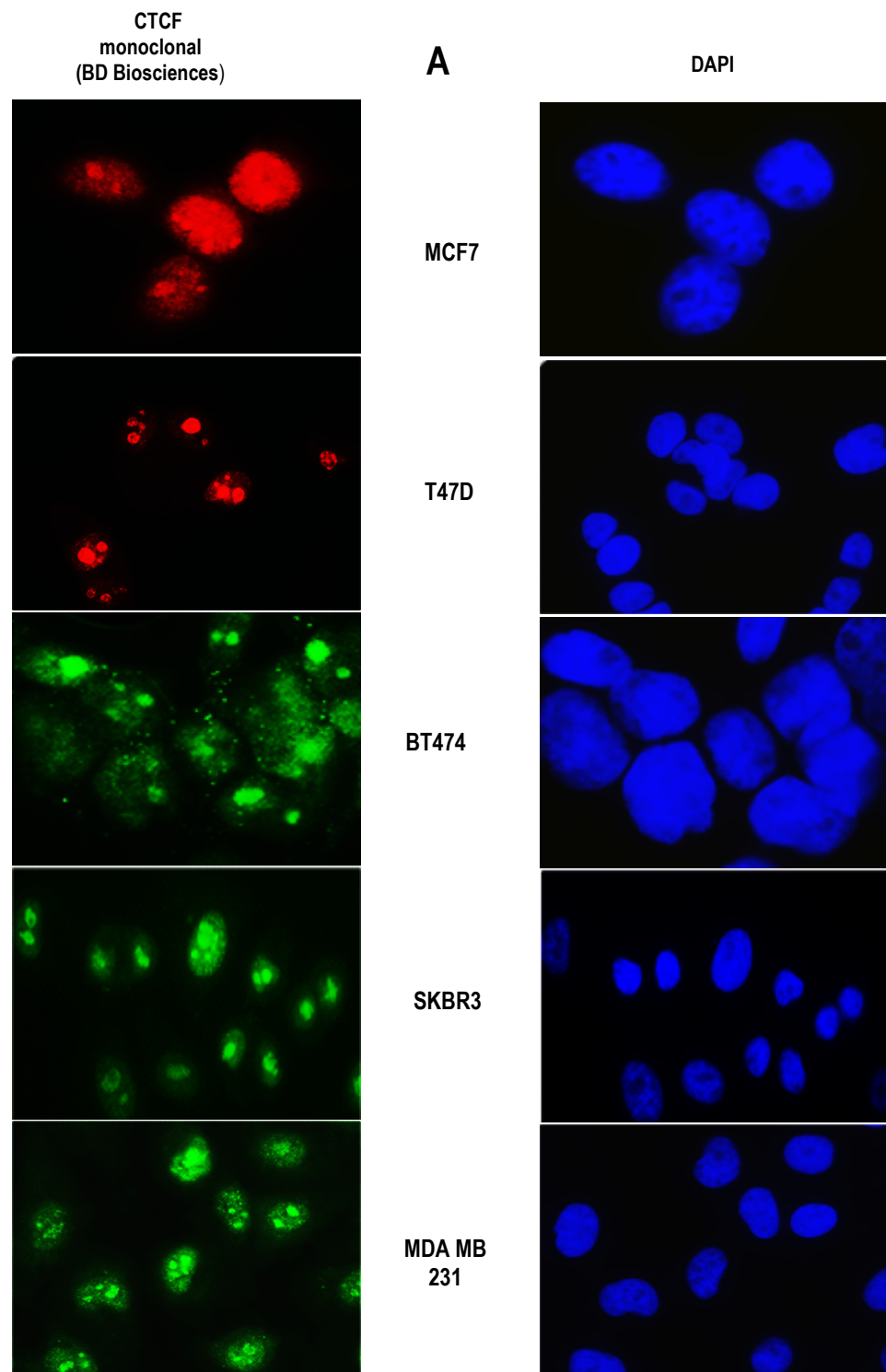
**Figure 3.3: Expression of proliferation markers Ki67 and PCNA in a panel of five different breast cancer cell lines and one normal breast epithelial cell line.** 50µg cell lysate of MCF7, T47D, BT474, SKBR3, MDA MB 231 breast cancer cells and the normal LDM 226 epithelial cell line cells were reduced on a 6.5% (for Ki67) and 8% (for PCNA) acrylamide gel for SDS PAGE and western blotting. Shown are uniform Ki67 expression bands (lower in T47D cells) detected with the monoclonal anti-Ki67 antibody from Vector but variable Ki67 expression across cell lines with the polyclonal antibodies from Abcam and Enzo.



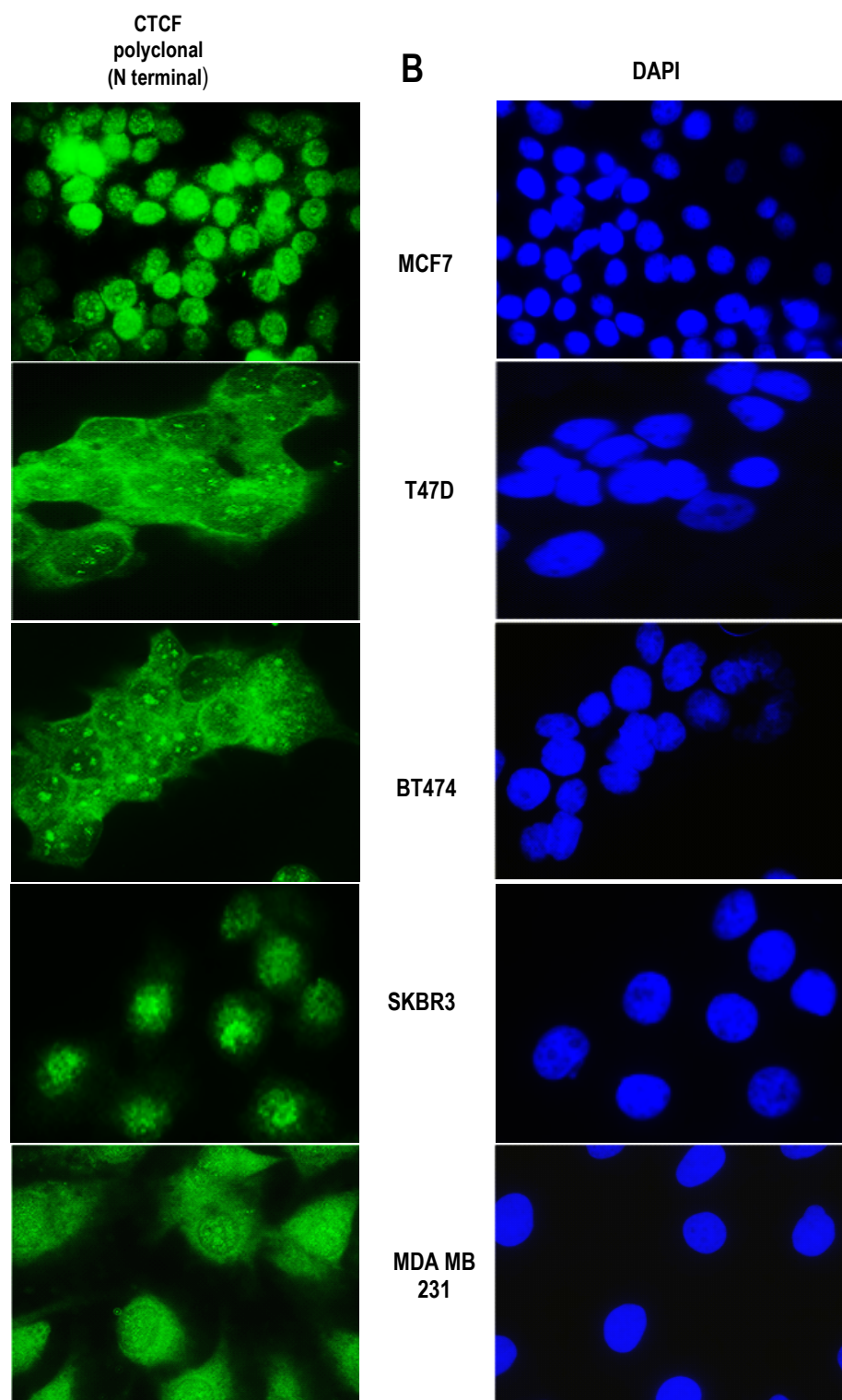
#### **3.4.4 CTCF localisation in relation to breast cancer phenotype and anti-CTCF antibody type**

Localisation and colocalisation studies are useful in hinting at the function of a protein through its cellular mapping with other proteins of known function (Bolte and Cordelieres, 2006). The nuclear location of CTCF in HeLa cells, breast cancer tissue and MCF7 cells is known (Zhang *et al.*, 2004; Rakha *et al.*, 2004; Torrano *et al.*, 2006) but it is not clear whether that localisation varies with breast cancer phenotype or the anti-CTCF antibody used to detect the protein. To establish whether there are any differences in CTCF localisation in breast cancer cell lines possessing different HR, HER2 and invasive profiles, the breast cancer cell lines used in this study underwent single immunofluorescence staining using different anti CTCF antibodies. The antibodies were the commercial product from BD Biosciences (also used in the western blotting of figure 3.2B) and N-terminal antibody (a kind gift from Prof Klenova of Essex University). The commercial antibody is marketed as detecting one CTCF isoform (140kDa), however as shown in figure 3.2B it may detect more isoforms by western blot. The in-house antibody is believed to detect two CTCF isoforms (180kDa and 130kDa) on western blotting (Docquier *et al.*, 2009). The primary antibodies were counterstained with an appropriate secondary antibody conjugated to TRITC (red) or FITC (green). Immunostaining with the monoclonal (BD Biosciences) anti CTCF antibody revealed a strong and predominantly nucleolar CTCF expression across all cell lines (figure 3.4A). The N-terminal polyclonal antibody showed a more diffuse, speckled nucleoplasmic expression and also some nucleolar expression across all cell lines (figure 3.4B). The MDA MB 231 cells revealed a completely diffuse nucleoplasmic expression with the polyclonal antibody (figure 3.4B, MDA MB 231). The diffuse nucleoplasmic expression of CTCF is supported by previous work in HeLa cells, breast cancer tissue and MCF7 cells (Zhang *et al.*, 2004; Rakha *et al.*, 2004; Torrano *et al.*, 2006). It is as yet not possible to confirm whether the nucleoplasmic CTCF distribution detected in figure 3.4 (A and B) is

from the 180kDa isoform as there is no specific antibody (commercial or in-house) able to identify the 180kDa isoform in immunofluorescence assays. Moreover, the N terminal anti-CTCF antibody went out of production and was not used in further experiments in this thesis. Taken together it would appear that different antibodies might detect CTCF to different extent in nucleoplasmic and nucleolar locations.



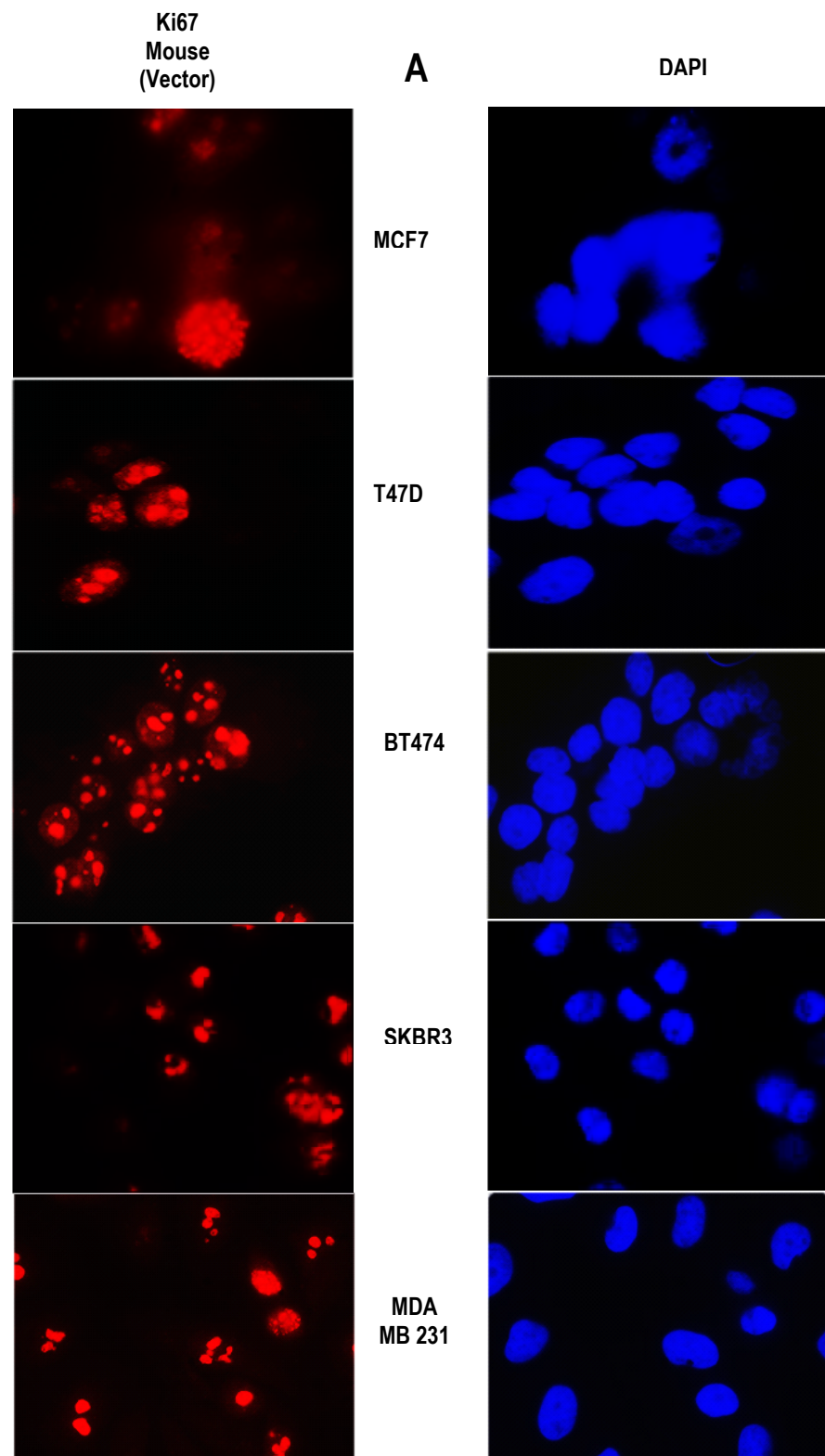
**Figure 3.4. Single indirect immunofluorescence staining of five breast cancer cell lines with CTCF primary antibodies.** (A) showing strong nucleolar localisation of CTCF protein in all cell lines, detected with the monoclonal antibody from BD Biosciences; and (B) indicating diffuse nucleoplasmic and nucleolar CTCF expression with the polyclonal anti-CTCF antibody.



**Figure 3.4 continued.**

### **3.4.5 Ki67 protein localisation in a panel of five breast cancer cell lines using two different anti Ki67 antibodies**

To confirm the localisation of Ki67 protein and determine possible differences in expression profile in a panel of five breast cancer cell lines with varying hormone receptor and invasive potential, single indirect immunofluorescence studies were performed. The experiment was carried out to also assess staining patterns of Ki67 protein using anti Ki67 antibodies derived from different animal species, as these antibodies would be used in further co-immunofluorescence studies. The five breast cancer cell lines were incubated with anti-Ki67 antibody (Vector VP K-452 [mouse monoclonal] and Abcam 833 [rabbit polyclonal]) at 1:200 dilutions respectively. They were counterstained with an appropriate secondary antibody conjugated to TRITC (red) or FITC (green). As shown in figure 3.5 (A – Ki67 monoclonal; and B – Ki67 polyclonal), there was a distinct nucleolar Ki67 protein expression with both the monoclonal and polyclonal antibodies and in all cell lines. In addition, MCF7 cells showed a perinucleolar (with the polyclonal antibody) and a punctate nucleoplasmic (with both antibodies) Ki67 expression profile. The nucleolar and nucleoplasmic distribution of Ki67 protein is supported by previous work (Bullwinkel *et al.*, 2006). The percentage Ki67 expression across all cell lines is shown in table 3.1. The two antibodies revealed markedly disparate Ki67 expression rates for T47D and BT474 cells. The relative expression percentages however are in general agreement with the western blot results in figures 3.3 A-C where Ki67 expression was shown to be variable across the cell lines. This would suggest that Ki67 expression was high in MCF7 cells (at variance with table 1.2) but low in both T47D (in agreement with table 1.2) and SKBR3 cell lines (at variance with table 1.2).



**Figure 3.5. Single immunofluorescence staining of five breast cancer cell lines with Ki67 primary antibodies.** Showing (A and B) strong nucleolar expression of Ki67 with both monoclonal and polyclonal anti Ki67 antibodies in all cell lines. MCF7 cells also showed nucleoplasmic Ki67 expression with the monoclonal antibody from Vector (A).

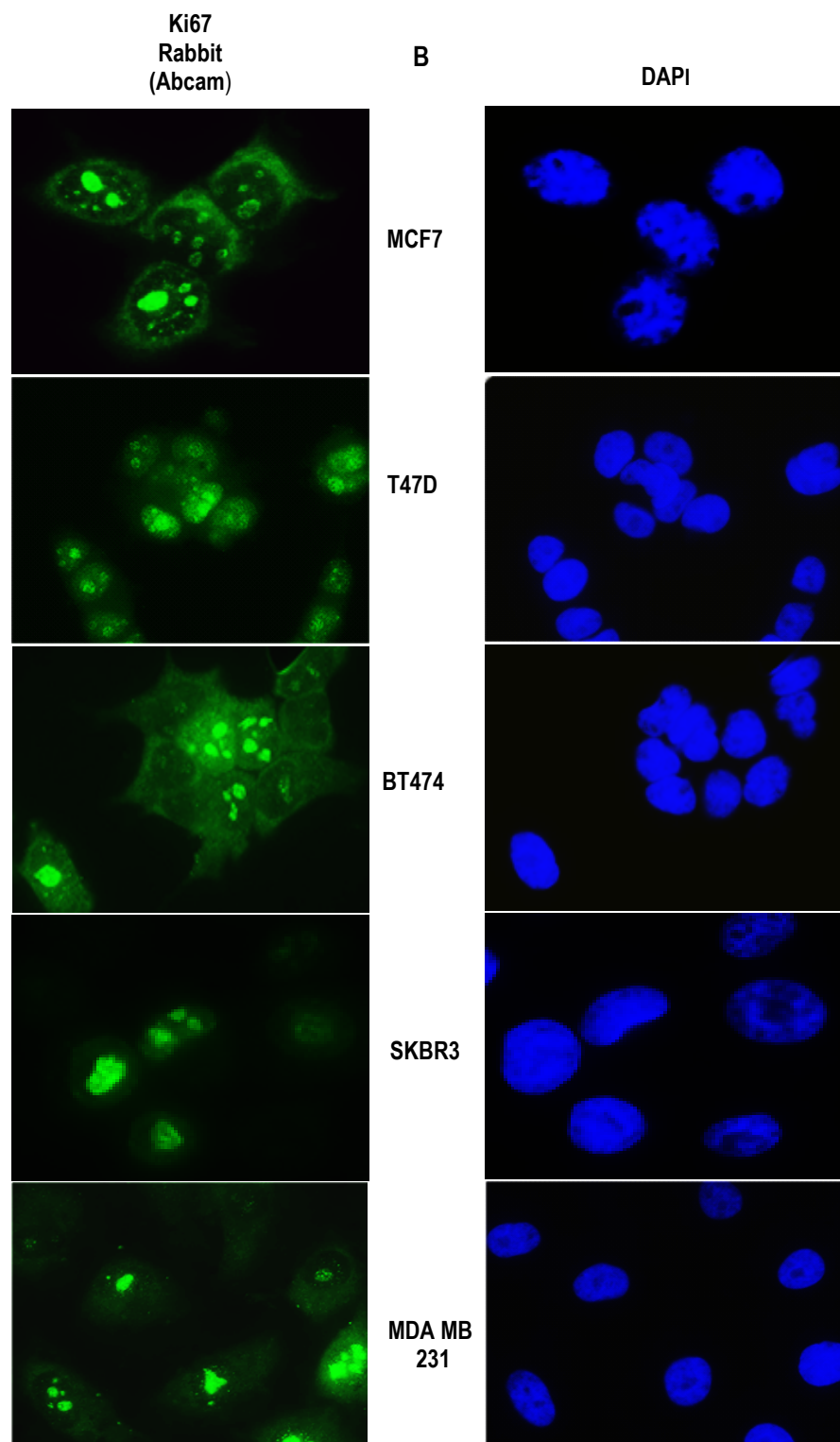


Figure 3.5 contd

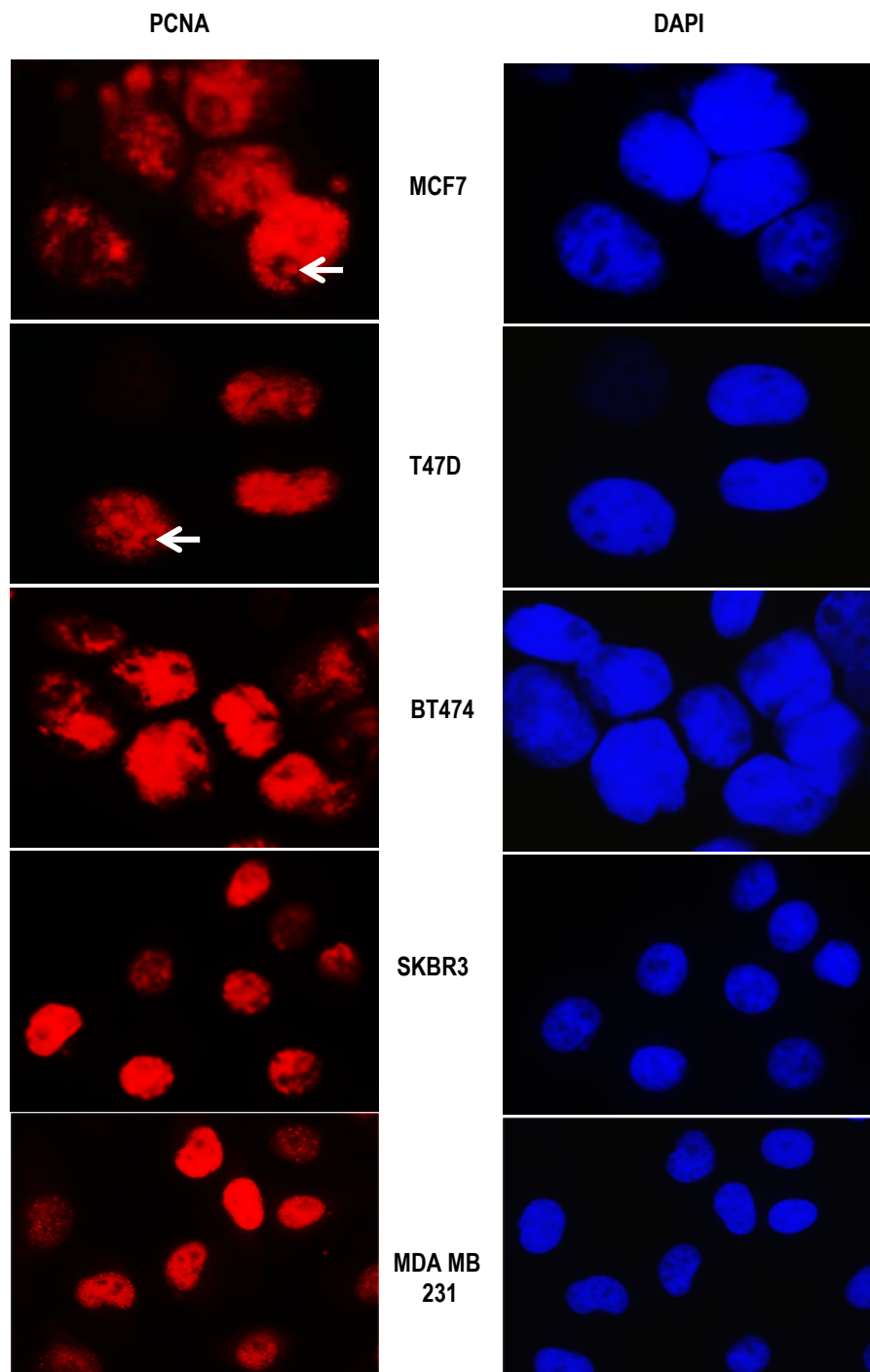
Breast cancer classification	Hormone receptor / HER2 expression	Breast cancer cell line	Ki67 expression level	
			From fig. 3.5 A and B	
			(Immunofluorescence)	
			Monoclonal	Polyclonal
Luminal A	ER+ PR+ HER2-	MCF7	80%	100%
		T47D	43%	90%
Luminal B	ER+ PR+ HER2+	BT474	76%	38%
HER2	ER- PR- HER2+	SKBR3	75%	50%
Claudin low	ER- PR- HER2-	MDA MB 231	90%	75%

**Table 3.1. Relative expression levels of Ki67 protein in five breast cancer cell lines.** Five breast cancer cell lines were subjected to single immunofluorescence using two different anti Ki67 antibodies. The percentage Ki67 expression was calculated by counting the number of cell nuclei that were stained with Ki67 relative to the total number of cells identified by DAPI in each field. The two anti Ki67 antibodies showed differential staining most evident for T47D and BT474 cell lines. The average percentages calculated from three different experiments are shown.



### **3.4.6 PCNA protein localisation localisation in breast cancer cell lines**

The proliferation marker PCNA functions as a DNA clamp and is involved in DNA transcription and DNA repair mechanisms (Moldovan *et al.*, 2007). It has been shown to reside in the nucleus by immunofluorescence in cells of the CV-1 cell line, a monkey kidney cell line and has different subnuclear localisation in transformed human amnion cells depending on the cell cycle stage (Celis and Celis, 1985; Waseem and Lane, 1990). To determine the pattern of PCNA expression in breast cancer cells with different invasive potential / hormone receptor status, single indirect immunofluorescence (IF) was performed. Cells were stained with an anti PCNA primary antibody and counterstained with secondary antibody conjugated to TRITC (red). As shown in figure 3.6, PCNA protein expression was diffusely nucleoplasmic in all cell lines. It also appeared to be nucleolar in the weakly invasive, hormone receptor positive MCF7 and T47D breast cancer cells as indicated by arrows in figure 3.6. These findings in breast cancer cell lines are in agreement with previous reports in other cell lines (Celis and Celis, 1985; Waseem and Lane, 1990). A summary of localisation patterns for CTCF, Ki67 and PCNA in all the breast cancer cell lines is shown in table 3.2.



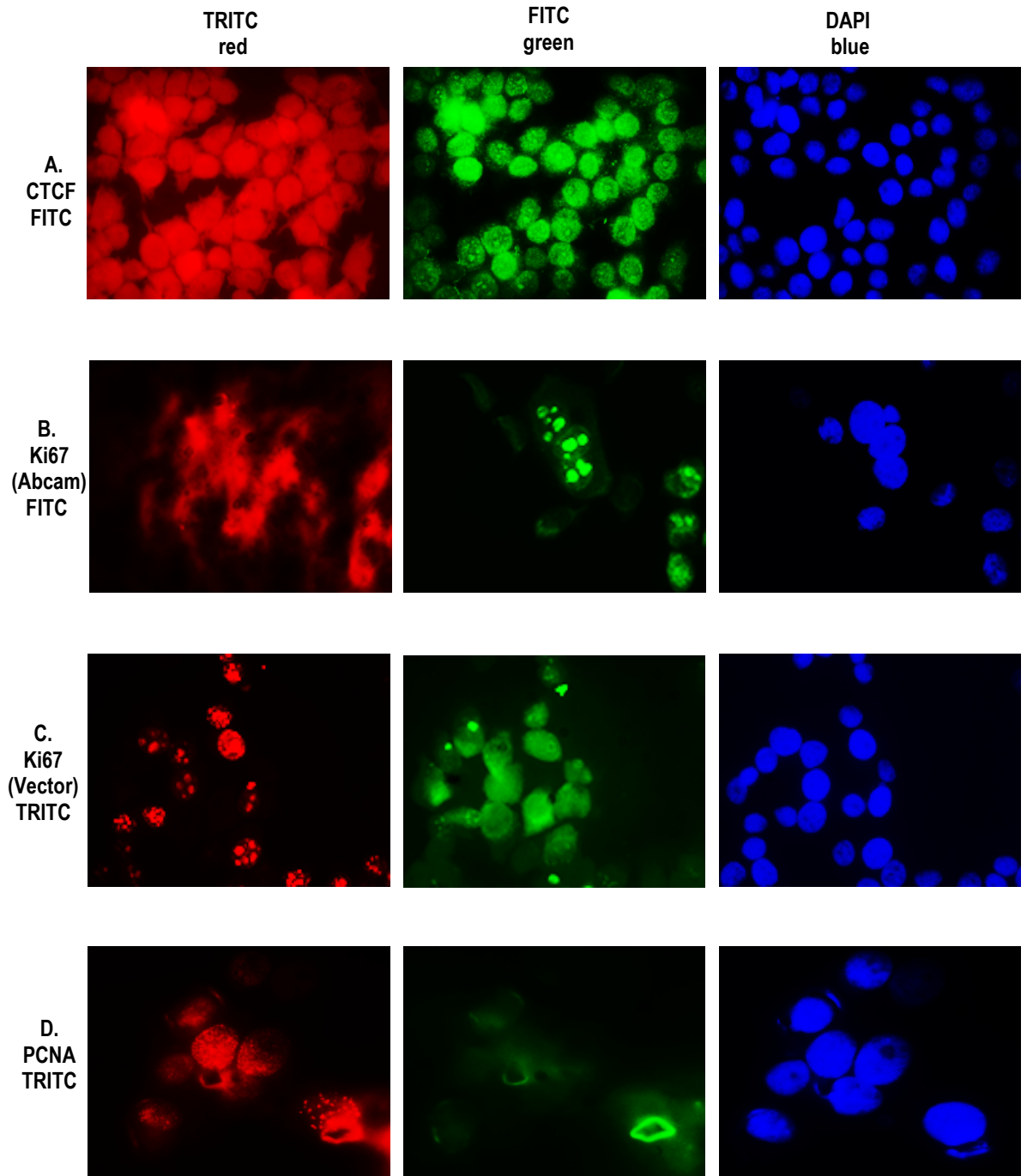
**Figure 3.6. Single immunofluorescence staining of five breast cancer cell lines with PCNA primary antibody.** Shown is the diffuse nucleoplasmic expression of PCNA in all cells lines. Also indicated by arrows is the nucleolar PCNA expression in the MCF7 and T47D cells.

	<b>MCF7</b>	<b>T47D</b>	<b>BT474</b>	<b>SkBr3</b>	<b>MDA MB 231</b>
	<b>ER+/PR+</b>	<b>ER+/PR+</b>	<b>ER+/PR+</b>	<b>ER-/PR-</b>	
	<b>HER2-</b>	<b>HER2-</b>	<b>HER2+</b>	<b>HER2+</b>	<b>ER-/PR- HER2-</b>
	<b>Non-invasive</b>	<b>Non-invasive</b>	<b>Moderately invasive</b>	<b>Moderately invasive</b>	<b>Aggressive</b>
<b>CTCF monoclonal</b>	Nucleolar +Minimally nucleoplasmic	Nucleolar +Minimally nucleoplas mic	Nucleolar +Minimally nucleoplasmic	Nucleolar +Minimally nucleoplasmic	Nucleolar +Minimally nucleoplas mic
<b>CTCF N terminal polyclonal</b>	Nucleoplasmic +minimally nucleolar	Nucleoplas mic + nucleolar	Nucleoplasmic +minimally nucleolar	Nucleoplasmic +minimally nucleolar	Nucleoplas mic
<b>Ki67 monoclonal</b>	Nucleolar + nucleoplasmic	Nucleolar	Nucleolar	Nucleolar	Nucleolar
<b>Ki67 polyclonal</b>	Nucleolar (perinucleolar	Nucleolar	Nucleolar	Nucleolar	Nucleolar
<b>PCNA monoclonal</b>	Nucleoplasmic + Nucleolar	Nucleoplas mic + Nucleolar	Nucleoplasmic	Nucleoplasmic	Nucleoplas mic

**Table 3.2. Tabular annotation of expression patterns of CTCF, Ki67 and PCNA proteins by single immunofluorescence in breast cancer cell lines.** Five breast cancer cell lines (MCF7, T47D, BT474, SKBR3 and MDA MB 231) possessing different hormone phenotype and invasive potential were subjected to single immunofluorescence using anti-CTCF, anti-Ki67 and anti-PCNA primary antibodies. The subnuclear distribution of each protein in each cell line is shown.

### 3.4.7 Assessment of immunofluorescence bleed through

Secondary antibodies display spectral overlaps and can bleed through into a different channel on microscopy. Bleed through could confound results obtained in two-colour co-localisation experiments. In order to confirm the absence of significant bleed through from respective fluorophore-conjugated secondary antibodies, single indirect immunofluorescence staining (IF) of MCF7 was performed. MCF7 cells were stained with mouse IgG1 monoclonal antibody to CTCF (BD Biosciences, 0.125 µg/ml, A), the rabbit polyclonal antibody to Ki67 (Abcam 833, 0.025 µg/µl, B), mouse monoclonal antibody to Ki67 (Vector, VP K-452, 1:100 dilution, C) and the mouse monoclonal antibody to PCNA (Abcam 29, 0.25µg/µl, D). They were respectively counterstained with goat anti-mouse secondary IgG1-specific antibody conjugated to FITC at (Southern Biotech, 5µg/µl, green colour), goat anti-rabbit secondary antibody conjugated to FITC (Abcam 6717, 5µg/µl, green colour), goat anti-mouse secondary antibody conjugated to TRITC (Southern Biotech, 5µg/µl, red colour) and goat anti-mouse IgG2a-specific secondary antibody conjugated to TRITC (Southern Biotech, 5µg/µl, red colour). DAPI was used to identify the nuclei at a concentration of 5µg/ml (blue colour). Cells were inspected via both red and green channels. As shown in figure 3.7A, B and D, the punctate, discrete, nucleolar or nucleoplasmic expression of proteins were identified only when viewed through the colour channel corresponding to the correct secondary antibody - fluorophore conjugate. There is however some bleedthrough for figure 3.7C wherein Ki67 tagged with TRITC secondary (red) was also visualised though minimally in the green channel. Given the low to no bleedthrough, immunofluorescence experiments demonstrated sufficient optimisation and results could be interpreted correctly.

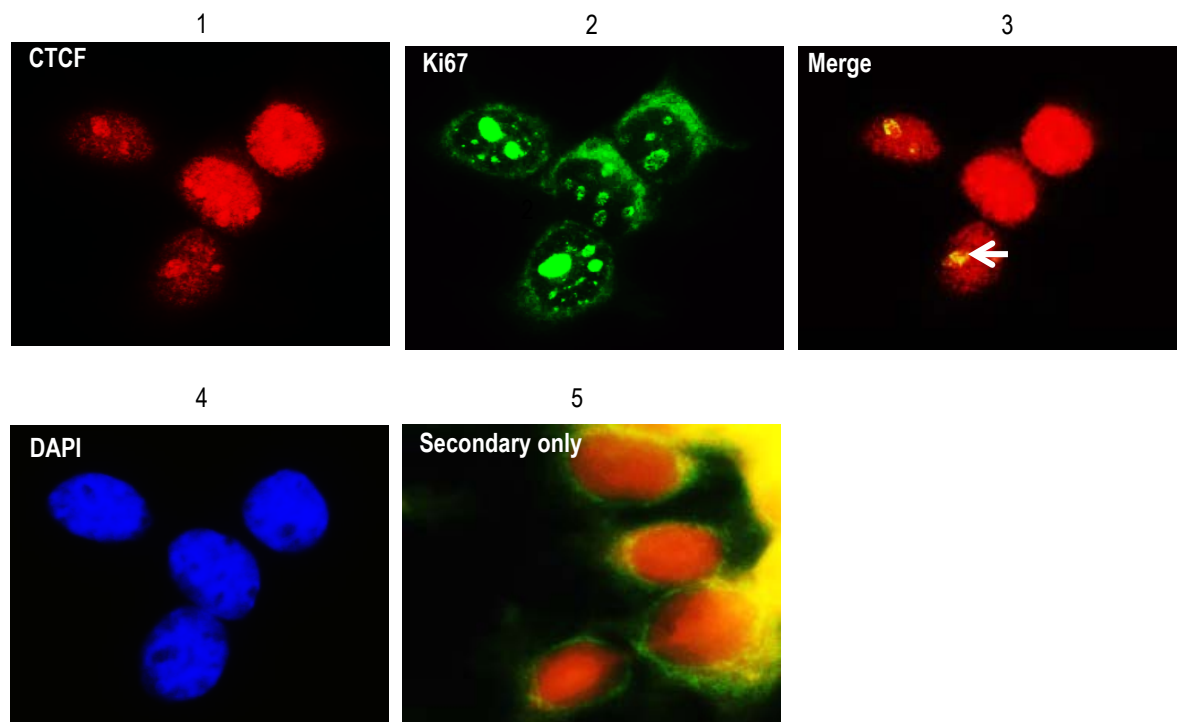


**Figure 3.7. Assessment of fluorophore-conjugated secondary antibody bleedthrough.** Showing no significant bleedthrough with anti CTCF antibody (A), polyclonal anti Ki67 antibody from Abcam (B), monoclonal anti Ki67 antibody from Vector (C) and the monoclonal anti PCNA antibody (D). There is minimal bleedthrough with the monoclonal anti Ki67 antibody from Vector (C).

### 3.4.8 CTCF protein co-localisation with Ki67 protein in breast cancer cell lines

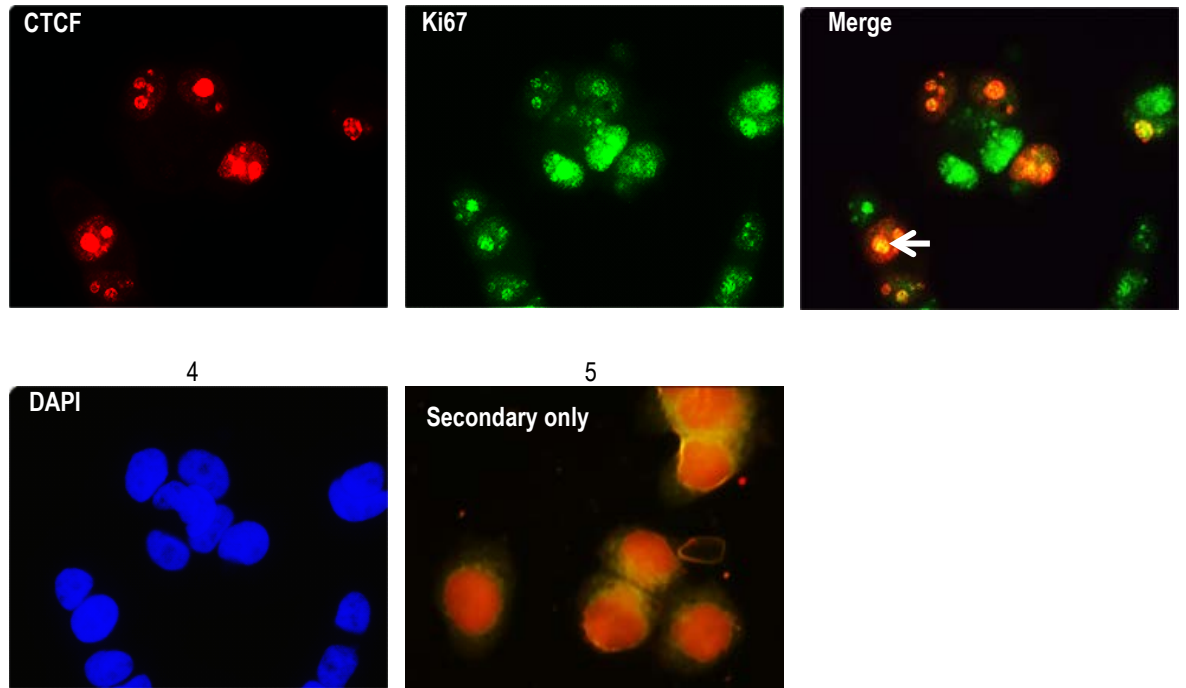
The presence of proteins in the same vicinity (colocalisation) in a cell compartment may suggest a functional link (Leclercq and Lacroix, 2004). To determine whether there was colocalisation of CTCF and Ki67 proteins in a panel of five breast cancer cell lines and to confirm the previously suggested colocalisation of these proteins (F. Docquier, personal communication), double indirect immunofluorescence staining of cells was performed. CTCF protein was identified with the mouse monoclonal anti-CTCF antibody (0.125 µg/ml) and goat anti-mouse secondary antibody Ig H+L conjugated to TRITC (Southern Biotech, 5µg/µl, red colour). Ki67 protein was detected with the rabbit polyclonal anti-Ki67 antibody (Abcam 833, 0.025 µg/µl) and goat anti-rabbit secondary antibody conjugated to FITC (Abcam 6717, 5µg/µl, green colour). DAPI (5µg/ml) was used to stain and identify the nuclei (blue colour). For control, cells were processed without incubation with primary antibody in order to rule out non-specific staining due to secondary antibodies. The results, seen in figure 3.8, revealed nucleolar colocalisation of CTCF and Ki67 proteins across all cell lines evidenced by the yellow colour change in the merge images (figures 3.8, 3, merge, arrowed). Some quiescent cells (identified by the absence of Ki67 expression) can be identified in BT474 and SKBR3 cells (figure 3.8 BT474C2; 3.8 SKBR3, D2). As these cells expressed CTCF (figures 3.8, BT474, C1; 3.8, SKBR3, D1) it might suggest that CTCF was significantly expressed in the quiescent state (G0 phase of the cell cycle) in these cell lines. This pattern of expression is not observed in all the cell lines studied and it could be speculated that it was a cell type – specific effect. Though these experiments were not designed to assess cytoplasmic expression of CTCF, the MDA MB 231 cells revealed no expression of CTCF protein in their well-defined cytoplasmic regions (figure 3.8, MDA MB 231). This finding agrees with that of Zhang *et al.* (2004) and Torrano *et al.* (2006) but differs from the cytoplasmic expression of CTCF in HeLa cells observed by Rakha *et al.* (2004) and supports the need for further investigation.

**A. MCF7 (ER+/PR+/HER2-)**

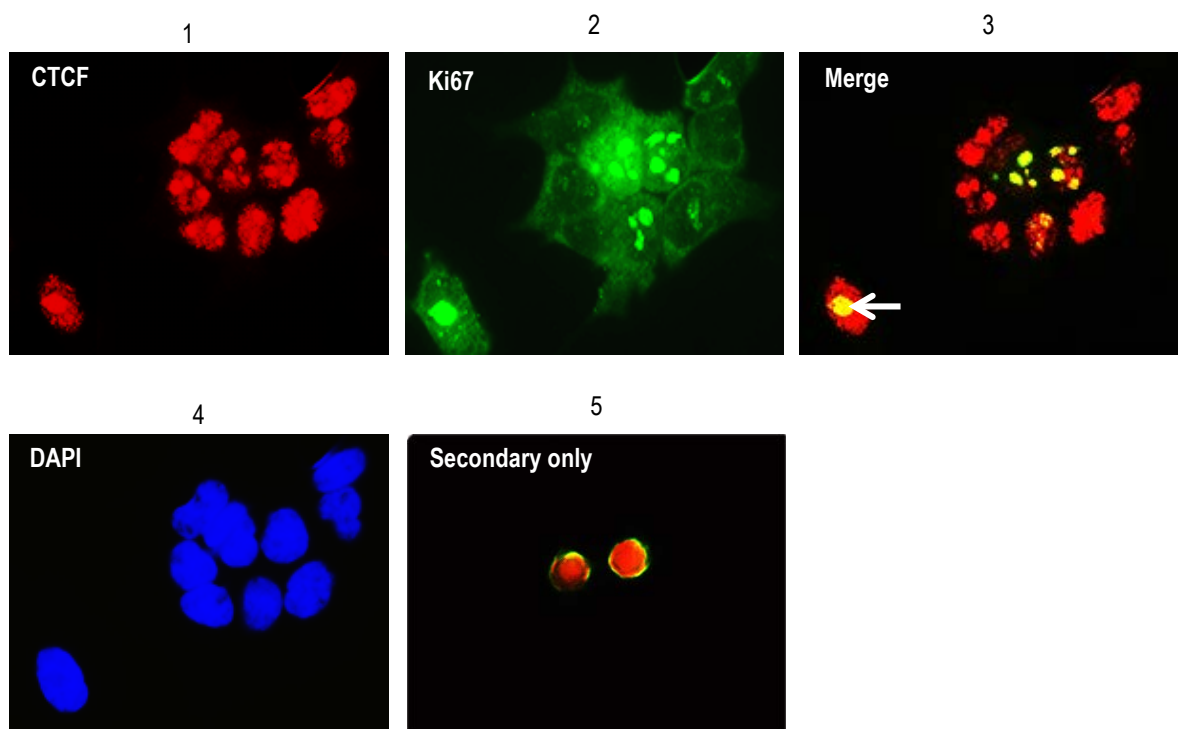


**Figure 3.8. A – E. Double immunofluorescence of CTCF protein with Ki67 protein in a panel of five breast cancer cell lines.** Showing mainly nucleolar CTCF (1) and Ki67 (2) protein expression. Colocalisation of CTCF and Ki67 proteins is indicated by an arrow in the merge images (3) across all cell lines.

**B. T47D (ER+/PR+/HER2-)**



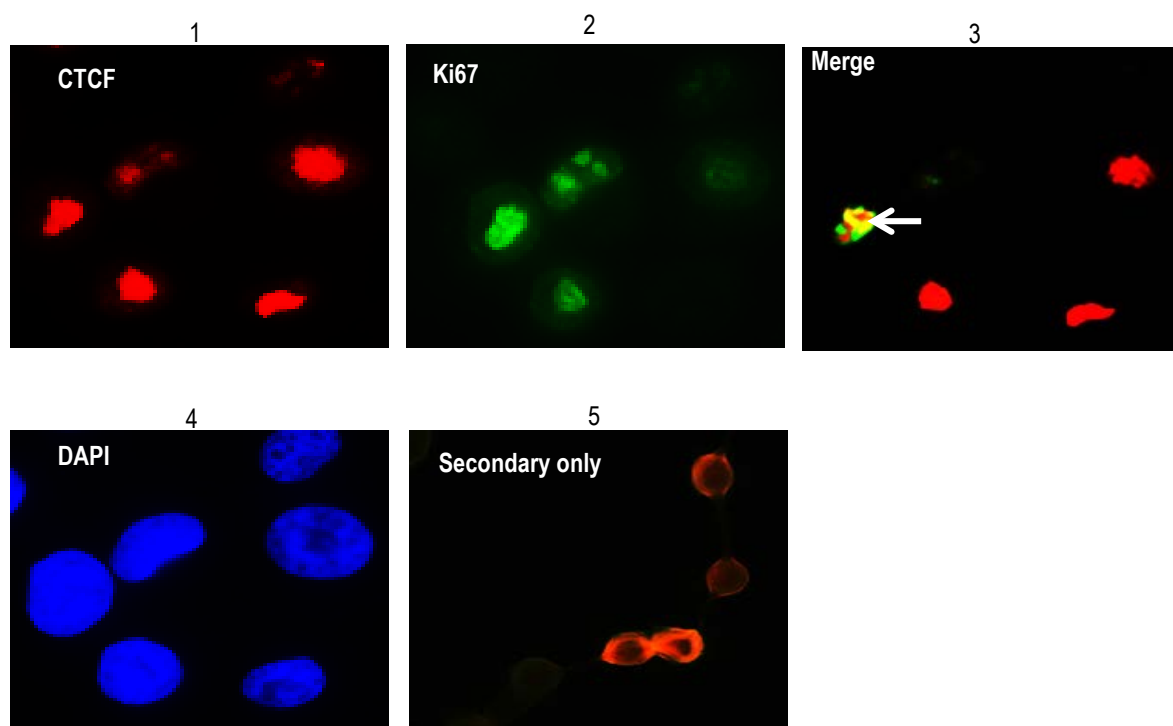
**C. BT474 (ER+/PR+/HER2+)**



**Figure 3.8 continued**



D. SKBR3 (ER-/PR-/HER2+)



E. MDA MB 231 (ER-/PR-/HER2-)

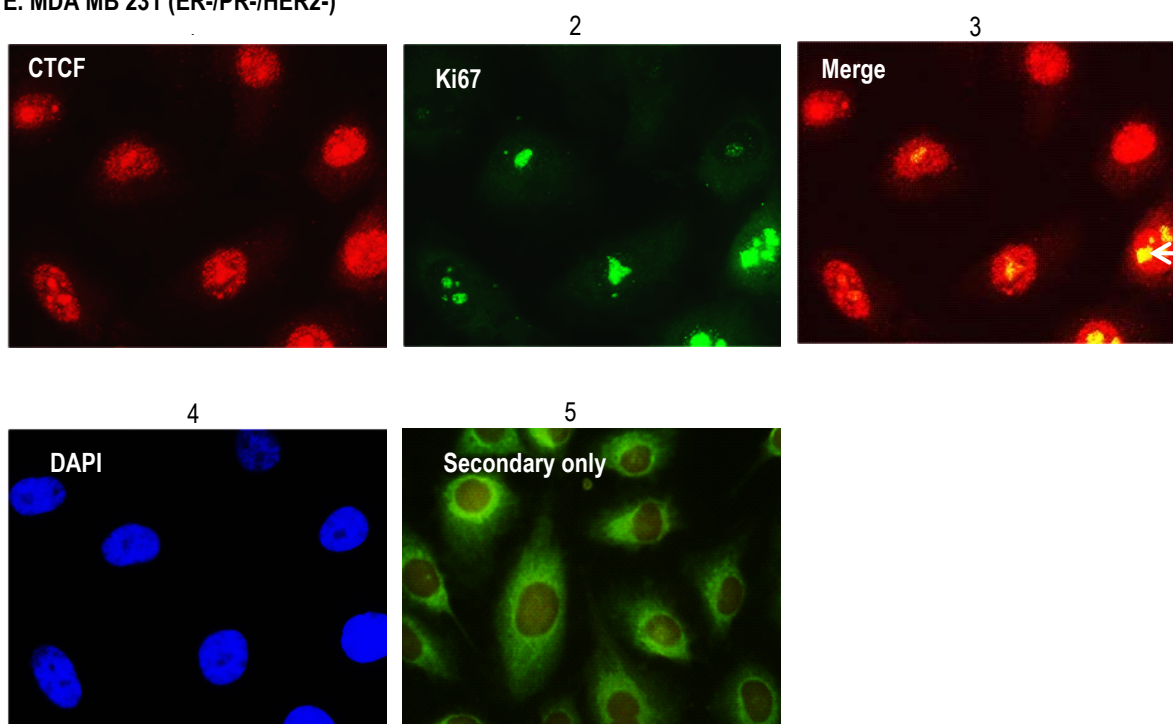
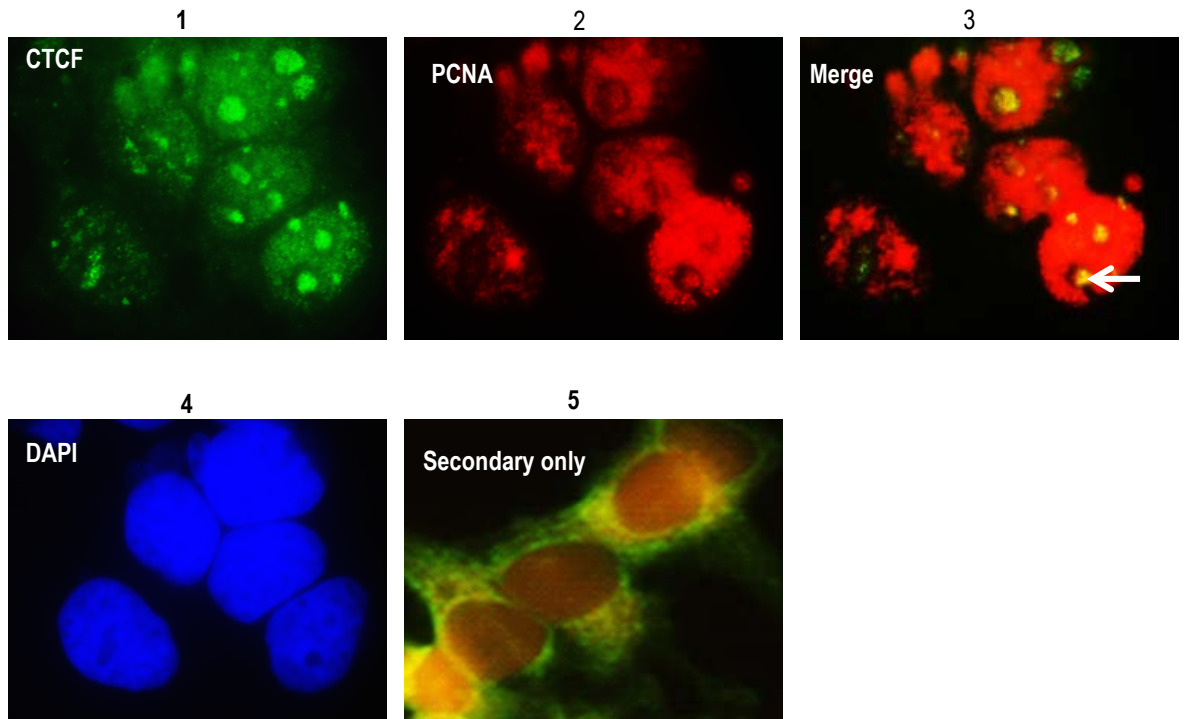


Figure 3.8 continued

### **3.4.9 CTCF protein colocalisation with PCNA protein in breast cancer cell lines with different hormone receptor / HER2 and invasive properties**

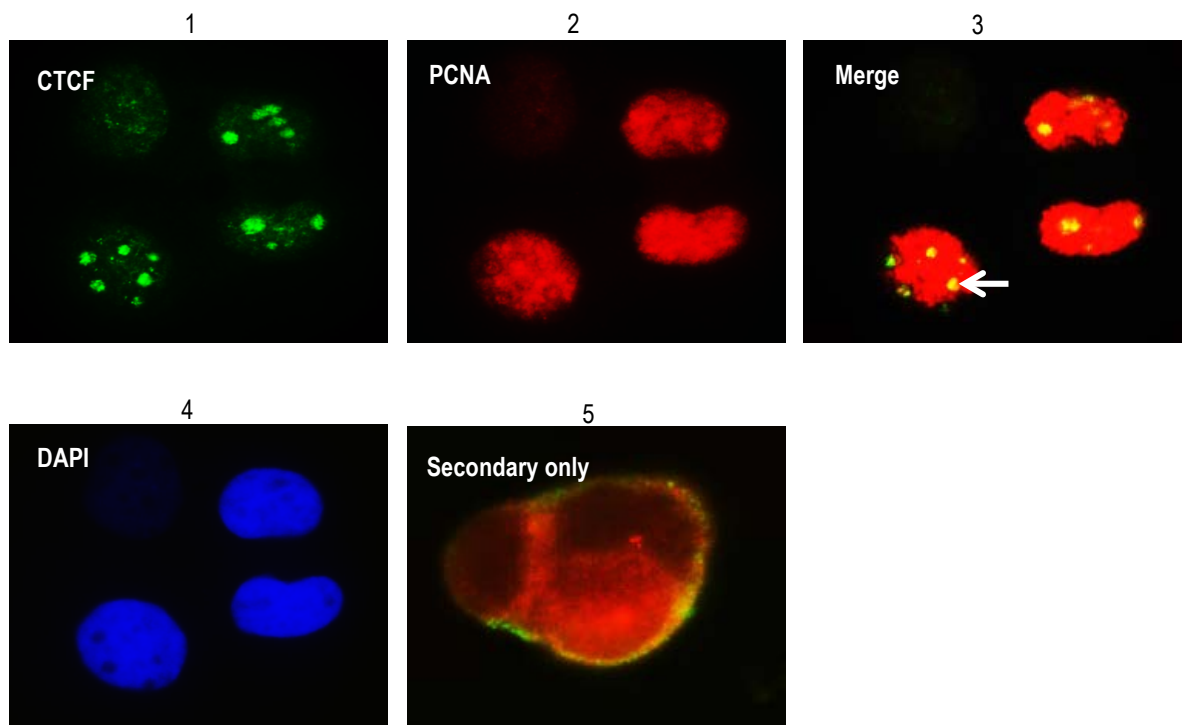
To confirm previously suggested colocation of CTCF and PCNA (in MCF7 cells) and to determine this colocalisation in other breast cancer cell lines with different HR / HER2 and invasive potential, double indirect immunofluorescence staining was performed with anti CTCF and anti PCNA antibodies. CTCF protein was identified with the mouse monoclonal anti-CTCF antibody (BD Biosciences, 0.125 µg/ml) and goat anti-mouse IgG1-specific secondary antibody conjugated to FITC (Southern Biotech, 5 µg/µl, green colour). PCNA protein was detected with the mouse monoclonal antibody (Abcam 29, 0.25µg/µl) and goat anti-mouse IgG2a-specific secondary antibody conjugated to TRITC (Southern Biochem, 5 µg/ml, red colour). DAPI (5µg/ml) was used to stain and identify the nuclei (blue colour). For control, cells were processed without incubation with primary antibody in order to rule out non-specific binding of secondary antibodies. CTCF expression was shown to be primarily nucleolar in all cell lines (figure 3.9, A –E, 1). PCNA expression appeared to be diffusely nucleoplasmic across all cell lines but was also nucleolar in MCF7 and T47D cells (figure 3.9, A – E, 2). Nucleolar colocalisation of CTCF and PCNA proteins was demonstrated in the minimally invasive, hormone receptor positive MCF7 and T47D cell lines (figures 3.9 A-3 and 3.9 B-3, merge, arrowed) but not in BT474, SKBR3 and MDA MB 231 cell lines. Though CTCF expression in MCF7 cells was also nucleoplasmic, there was no colocalisation with nucleoplasmic PCNA protein whose expression in that compartment was extensive. The colocalisation patterns of CTCF with Ki67 and PCNA is summarised in table 3.3.

**A. MCF7 (ER+ / PR+ / HER2-)**

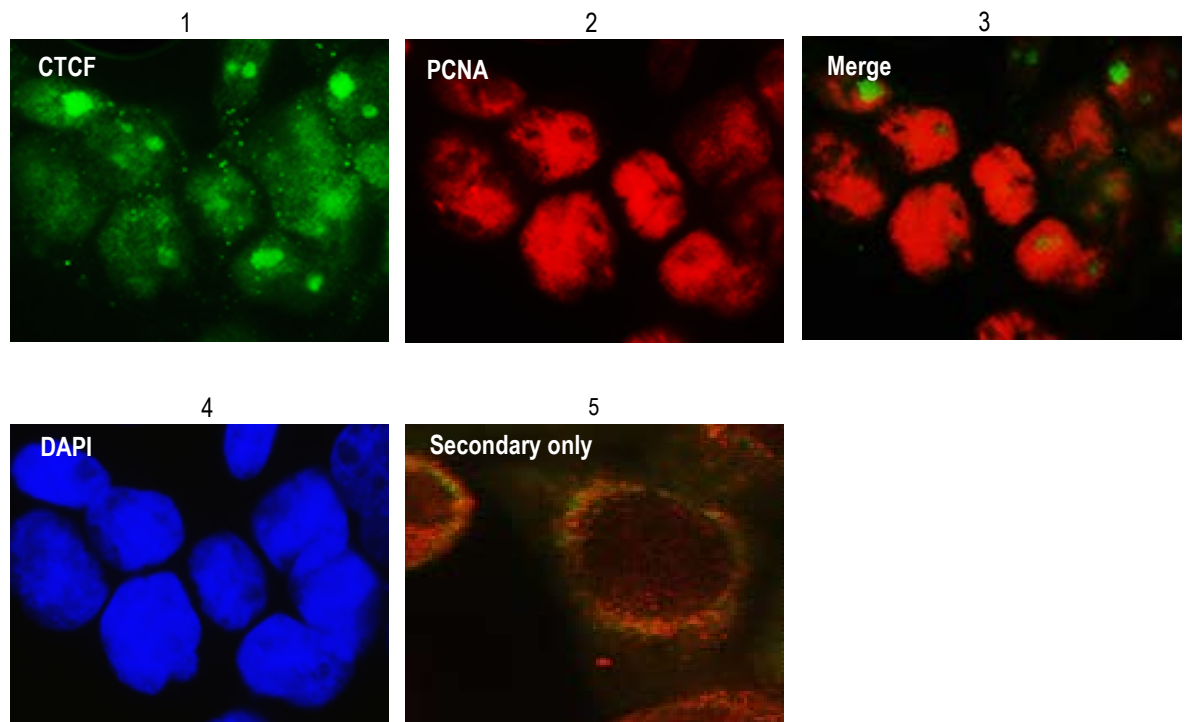


**Figure 3.9 A - E. Double indirect immunofluorescence of CTCF with PCNA in a panel of five breast cancer cell lines.** Demonstrated is the predominant nucleolar expression of CTCF in all cell lines (CTCF panel, 1). Extensive nucleoplasmic expression of PCNA is shown in all cell lines (PCNA panel, 2). In addition, A (MCF7) and B (T47D), revealed nucleolar PCNA presence. Colocalisation of CTCF and PCNA proteins is indicated by an arrow in the merge images (3) in A and B.

**B. T47D (ER+/PR+/HER2-)**

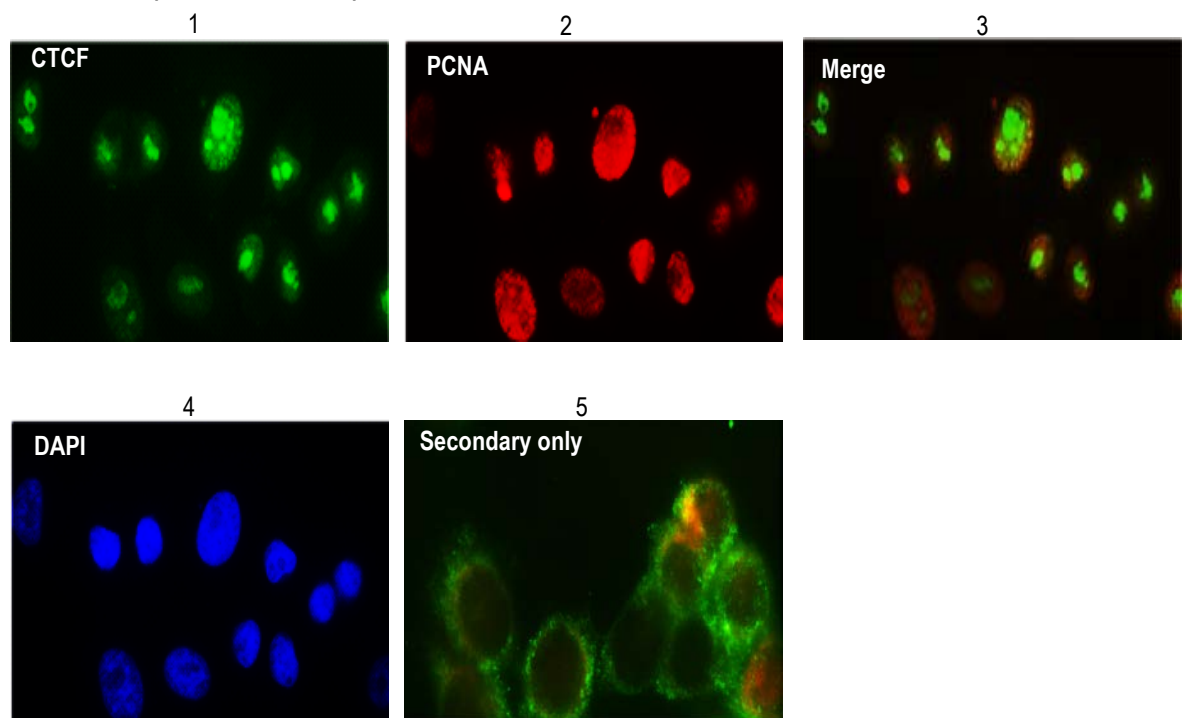


**C. BT474 (ER+/PR+/HER2+)**

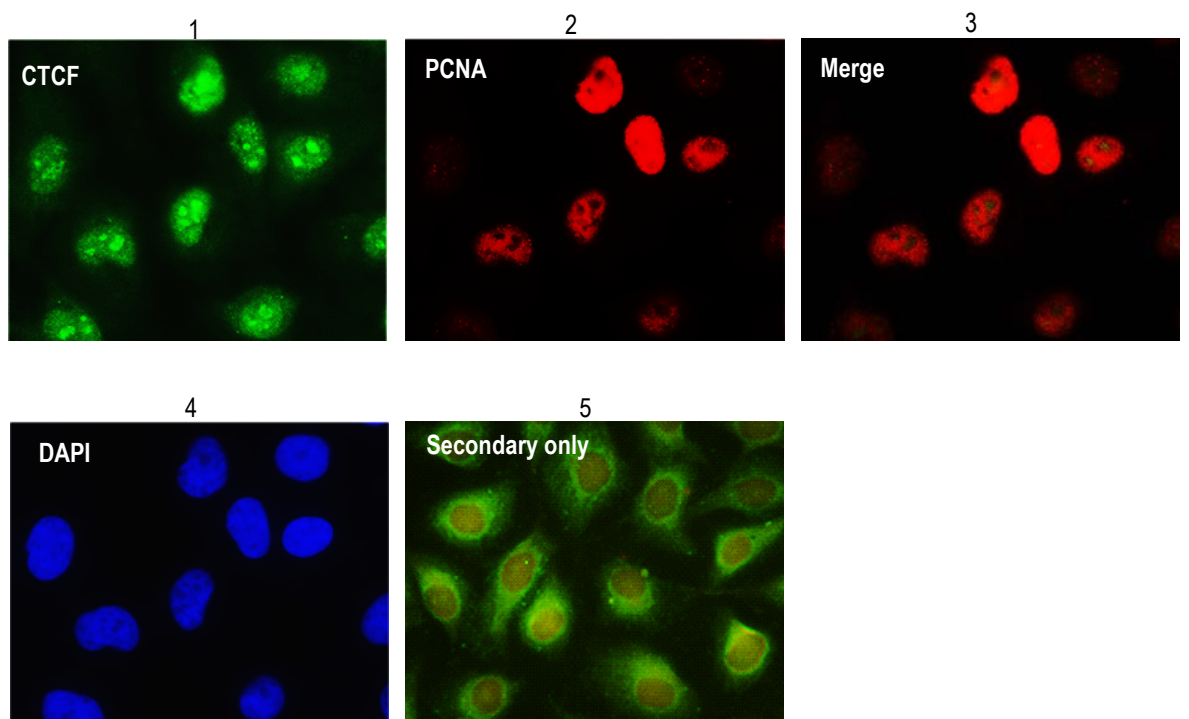


**Figure 3.9 continued**

**D. SKBR3 (ER-/PR-/HER2+)**



**E. MDA-MB-231 (ER-/PR-/HER2-)**



**Figure 3.9 continued**

	<b>MCF7</b>	<b>T47D</b>	<b>BT474</b>	<b>SkBr3</b>	<b>MDA MB 231</b>
	<b>(ER+/PR+)</b>	<b>(ER+/PR+)</b>	<b>(ER+/PR+)</b>	<b>(ER-/PR-)</b>	<b>(ER- / PR-)</b>
	<b>HER2-</b>	<b>HER2-</b>	<b>HER2+</b>	<b>HER2+</b>	<b>HER2-</b>
	<b>Non-invasive</b>	<b>Non-invasive</b>	<b>Moderately invasive</b>	<b>Moderately Invasive</b>	<b>Aggressive</b>
<b>CTCF / Ki67</b>	CTCF- nucleolar Ki67- nucleolar <b>Colocalisation</b>	CTCF- nucleolar Ki67- nucleolar <b>Colocalisation</b>	CTCF- nucleolar Ki67 - nucleolar <b>Colocalisation</b>	CTCF- nucleolar Ki67- nucleolar <b>Colocalisation</b>	CTCF- nucleolar Ki67-nucleolar <b>Colocalisation</b>
<b>CTCF / PCNA</b>	CTCF- nucleolar PCNA- nucleoplasmic, nucleolar <b>Colocalisation</b>	CTCF - nucleolar PCNA - nucleoplasmic, nucleolar <b>Colocalisation</b>	CTCF- nucleolar PCNA- nucleoplasmic, diffuse <b>No colocalisation</b>	CTCF- nucleolar PCNA- nucleoplasmic <b>No colocalisation</b>	CTCF- nucleolar PCNA- nucleoplasmic <b>No colocalisation</b>

**Table 3.3. Summary of colocalisation of CTCF with Ki67 and PCNA in a panel of five breast cancer cell lines.** Tabulated summary of colocalisation patterns between CTCF and Ki67 / PCNA proteins in a panel of five breast cancer cell lines. CTCF and Ki67 colocalised in the nucleolus of all five breast cancer cell lines. CTCF and PCNA colocalised only in the hormone receptor positive, weakly invasive MCF7 and T47D breast cancer cell lines.

#### **3.4.10 CTCF immunoprecipitation (IP) and coimmunoprecipitation (co-IP) with Ki67 and PCNA in MCF7 breast cancer cells using a high stringency IP buffer**

Protein colocalisation suggests that colocalising proteins could be functionally linked (Leclercq and Lacroix, 2004). In order to assess whether colocalising proteins observed in Figures 3.8 and 3.9 were physically bound in a functional complex, CTCF was immunoprecipitated from MCF7 cells and then probed for Ki67 and PCNA co-immunoprecipitation. A successful protein immunoprecipitation experiment is critically dependent on the IP lysis buffer solution (Klenova *et al.*, 2002). As CTCF immunoprecipitation experiments in breast cancer cells have not previously been optimised, a high stringency IP lysis buffer was therefore first used to perform CTCF IP experiments. CTCF immunoprecipitation from MCF7 cell lysate was performed using 20µl (20µg – 60µg) of anti CTCF antibody (Millipore, monoclonal) and a high stringency buffer (buffer 1) containing Tween-20 as detergent and a high salt concentration (see appendix section 8.2 for composition). An equivalent concentration of IgG antibody was used for IP in MCF7 cells and served as non-specific IP control. To determine the efficiency of the immunoprecipitation process, silver staining of an SDS PAGE resolved gel was performed (figure 3.10A). The silver stained gel revealed a smear of protein bands in the input lane as expected in a complex mixture of proteins in a cell lysate (figure 3.10A). The precleared lysate lane revealed a less dense smear of protein suggesting clearing of nonspecific protein binding from the input by the beads. There are distinct protein bands displayed in the eluted CTCF IP lane that are quite different from the control eluted IgG IP lane.

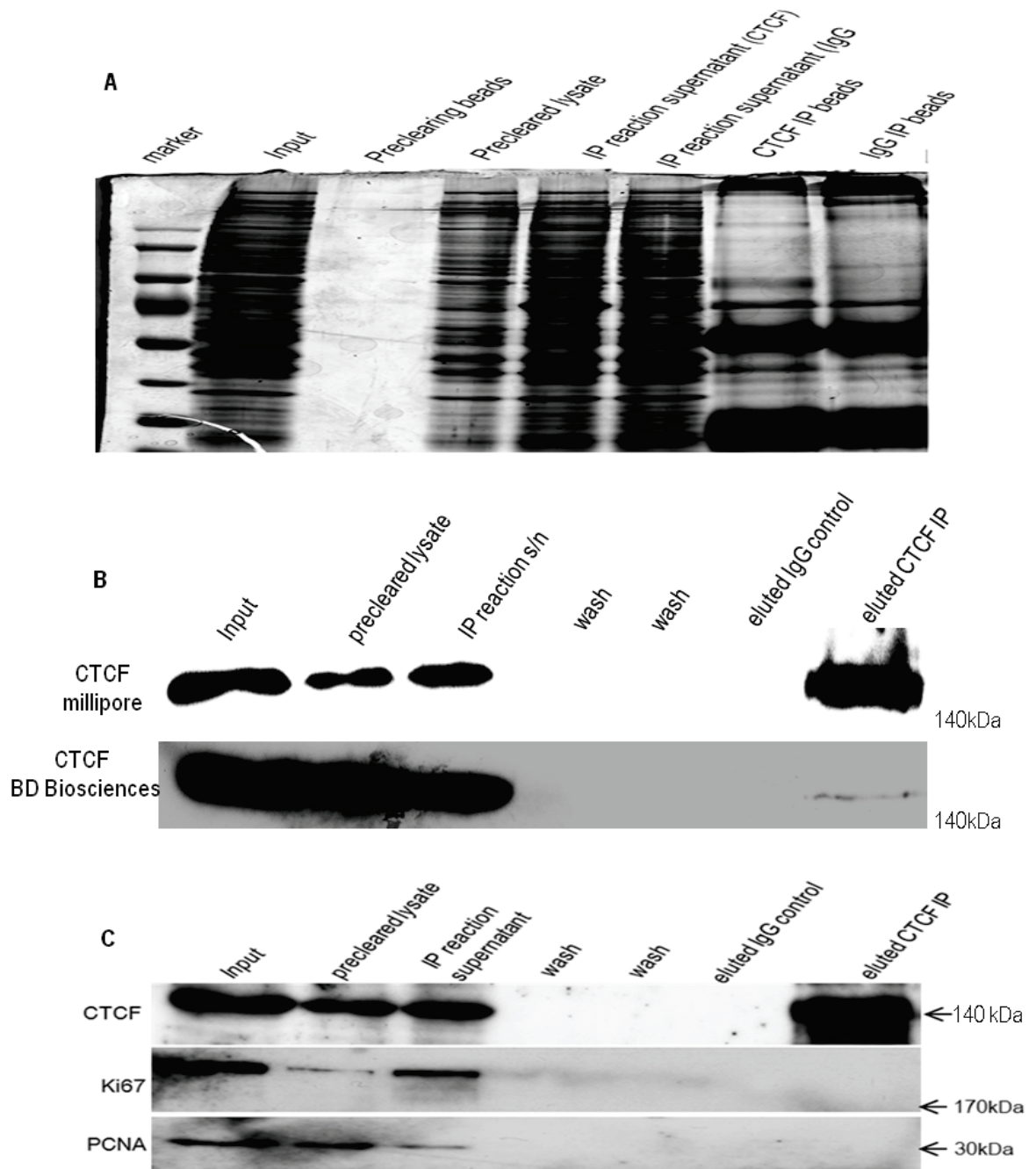
To confirm the specificity of the interaction, two CTCF antibodies raised from different species recognising different epitopes were used to assess CTCF IP. The blot of the CTCF IP was therefore probed with the rabbit monoclonal antibody to CTCF (Millipore, 1:5000 dilution) and the mouse monoclonal antibody to CTCF (0.5 µg/ml; BD

Biosciences). As shown in figure 3.10B, both antibodies detected the immunoprecipitated (IPed) CTCF protein though to varying extent.

The blot of the CTCF IP reaction shown in figure 3.10C (top panel), revealed a strong protein band in the input. This was obtained with 5% (20µl) of total cell lysate volume (400µl). There was a decrease in this expression level in the precleared lysate lane in agreement with Figure 3.11A (~5.7% of lysate volume used: 20µl of 350µl). It may not be unexpected to observe a relatively higher protein level in the IP reaction supernatant lane as ~7% volume (20µl of 300µl IP reaction supernatant lysate volume) was loaded for SDS PAGE. In the CTCF IP elution lane a strong band of CTCF is shown to be successfully IPed, an expression almost double that of the input. About 16.6% of the IP eluate (10µl of 60µl total elution volume) was used for western blot analysis.

On assessing for coimmunoprecipitation by probing the blotted CTCF IP membrane sequentially with mouse monoclonal antibody to Ki67 (Vector, V K452, 1:1000 dilution) and mouse monoclonal antibody to PCNA (Ab29, 1:3000 dilution), the pattern of Ki67 and PCNA protein expression in the control lanes was noted to be similar to CTCF expression (figure 3.10C, middle and bottom panels). There was however no co-IP with Ki67 or PCNA demonstrated as no protein expression band was observed in the 'eluted CTCF IP' lane (figure 3.10C, middle and bottom panels). The IgG IP negative control lane as expected demonstrated no protein bands.





**Figure 3.10. Silver stain (A), CTCF immunoprecipitation (B) and CTCF co-immunoprecipitation (C) in MCF7 cells using a high stringency buffer.** Shown in the silver stain (A) is the smear of protein bands in the input and distinct protein bands in CTCF elution lanes. (B) Demonstrated CTCF immunoprecipitation using anti CTCF antibodies raised from rabbit (Millipore) and mouse (BD Biosciences) confirming the specificity of the interaction. The band corresponding to the immunoprecipitated CTCF-140 isoform is indicated by the band in the 'eluted CTCF lane. (C) Revealed a lack of coimmunoprecipitation with Ki67 and PCNA evidenced by the absence of a protein band in the CTCF elution lanes in the Ki67 and PCNA panels.

#### **3.4.11 CTCF co-immunoprecipitation with Ki67, PCNA and known protein partners (RNA pol II and PARP 1) in MCF7 breast cancer cells using a medium stringency buffer.**

The lack of CTCF co-IP with Ki67 and PCNA shown in figure 3.10 might be due to the high stringency of the IP lysis buffer and / or a non-optimal experimental process (including number of washes in the IP process). To further determine whether Ki67 / PCNA could co-IP with CTCF IP and to assess the adequacy of the experimental set up, a medium strength IP lysis buffer with less stringency was used and further CTCF co-IP with known partners, RNA pol II and PARP 1 also performed (Chernukhin *et al.*, 2007; Farrar *et al.*, 2010). CTCF protein was immunoprecipitated in MCF7 cell lysate using 20µl (20µg – 60µg) of CTCF (Millipore) antibody and a medium stringency buffer (buffer 2) containing NP-40 as detergent and a medium salt concentration (see appendix 8.2 for composition). An equivalent concentration of irrelevant IgG antibody was used to perform IP on MCF7 cells and served as negative control. As shown in figure 3.11A, CTCF protein expression decreased progressively as expected in the control lanes and is successfully IPed in the eluted IP beads lane. The IgG control was negative. To assess coimmunoprecipitation, the blotted membrane was serially probed with mouse monoclonal antibody to Ki67 (Vector, VP K452, 1:1000 dilution, B), mouse monoclonal antibody to PCNA (Ab29, 0.3 µg/µl, C), rabbit polyclonal antibody to RNA pol II (sc 899, 0.2 µg/ml, D), and mouse monoclonal antibody to PARP 1 (Enzo, 1:2000 dilution, E). There was no co-IP demonstrated with Ki67 and PCNA (figures 3.11 B and C). There was also no co-IP with known partners RNA pol II and PARP1 demonstrated (figure 3.11 D and E).

IP experiments performed with two anti Ki67 antibodies (Vector VP K452 and Abcam 833) using the same buffer composition and in MCF7 cell lysates did not demonstrate Ki67 immunoprecipitation (data not shown).



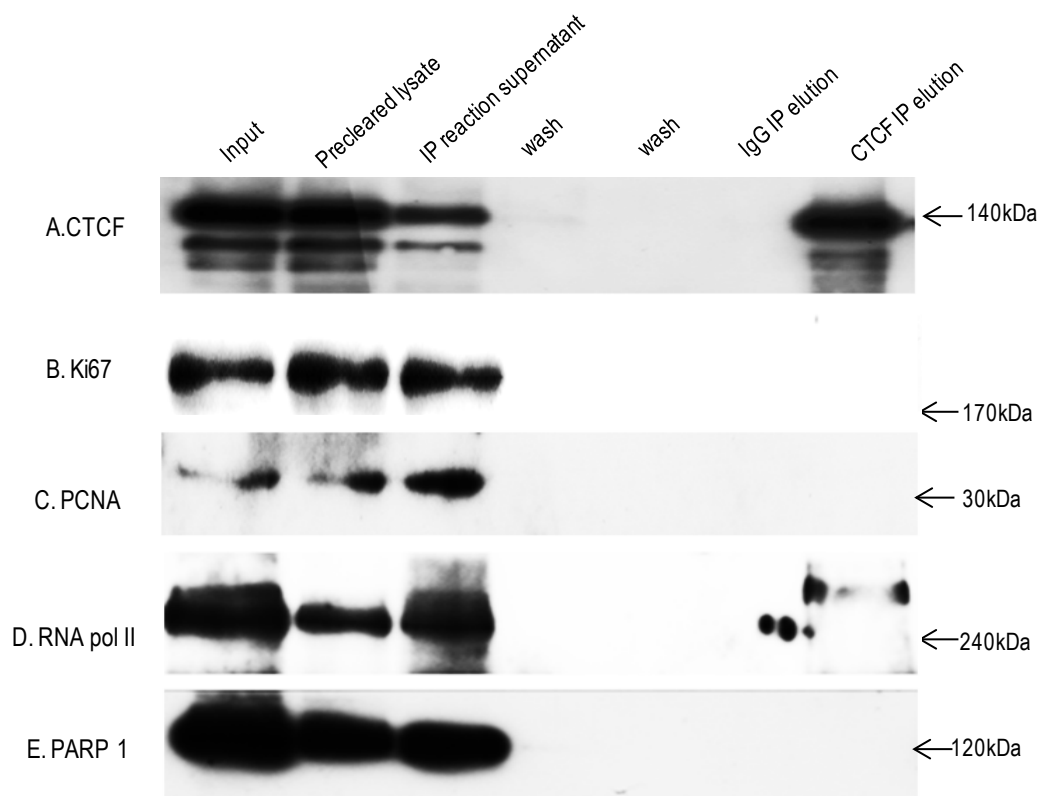
**Figure 3.11. CTCF co-IP with Ki67, PCNA, RNA pol II, and PARP 1 in MCF7 breast cancer cells using a medium stringency IP buffer.** Shown in the Figure is immunoprecipitated CTCF in the 'CTCF IP elution lane (A). While Ki67, PCNA, RNA pol II and PARP1 were all detected in the control lanes, there was no coimmunoprecipitation demonstrated in the elution lane as no protein band bands were observed in that lane (B - E).

#### **3.4.12. CTCF immunoprecipitation and coprecipitation with Ki67, PCNA and known protein partners (RNA pol II and PARP 1) in MCF7 breast cancer cells using a low stringency buffer.**

The continuing inability to co-IP Ki67, PCNA and known CTCF partners, RNA pol II and PARP1 shown in Figure 3.11, may suggest that the IP process itself was still not fully optimised. To increase the ability of CTCF to pull down its protein partners, further experiments were performed using a low stringency buffer. This buffer, while denaturing and able to release nuclear protein would not disrupt native protein conformation and would have minimal denaturing effect on antibody binding sites and therefore better preserve CTCF native interactions. CTCF protein was therefore immunoprecipitated in MCF7 cell lysates using 20 $\mu$ l (20  $\mu$ g – 60 $\mu$ g) of CTCF (Millipore) antibody and a low stringency IP lysis buffer containing NP-40 as detergent and a low salt concentration (buffer 3). An equivalent concentration of IgG antibody was used for IP in MCF7 cells and served as negative control. As shown in figure 3.12 A, B and C, CTCF was successfully IPed with mouse monoclonal antibody to CTCF (BD Biosciences, 0.5  $\mu$ g/ml, A) but there was no co-IP demonstrated on probing the blot membrane with mouse monoclonal antibody to Ki67 (Vector, V P K452, 1:1000 dilution, B) and mouse monoclonal antibody to PCNA (Ab29, 0.3  $\mu$ g/ $\mu$ l, C). Further assessment for coimmunoprecipitation with the rabbit polyclonal antibody to RNA pol II (sc 899, 0.2  $\mu$ g/ml) however, demonstrated a protein band coprecipitated in the IP elution lane as shown in figure 3.12 D. This band of protein is noted to be heavier than the bands detected in the input, precleared and IP reaction supernatant control lanes. It is quite possible that on account of hyperphosphorylation, complexing of CTCF with a pre-peptide form of RNA pol II or indeed other post translational modifications, the size of RNA pol II could be somewhat different as shown in this result. To determine whether this result was spurious or not, reverse IP was performed by immunoprecipitating RNA pol II protein in MCF7 cell lysates using 20 $\mu$ g of RNA pol II antibody. Though RNA pol II was successfully IPed, there was no co-IP demonstrated with CTCF (data not shown).

With regards to the results in Figure 3.12, further probing of the IP blot membrane with mouse monoclonal antibody to PARP 1 (Enzo, 1:2000 dilution, E) did not demonstrate coimmunoprecipitation with PARP1 (figure 3.12 E).

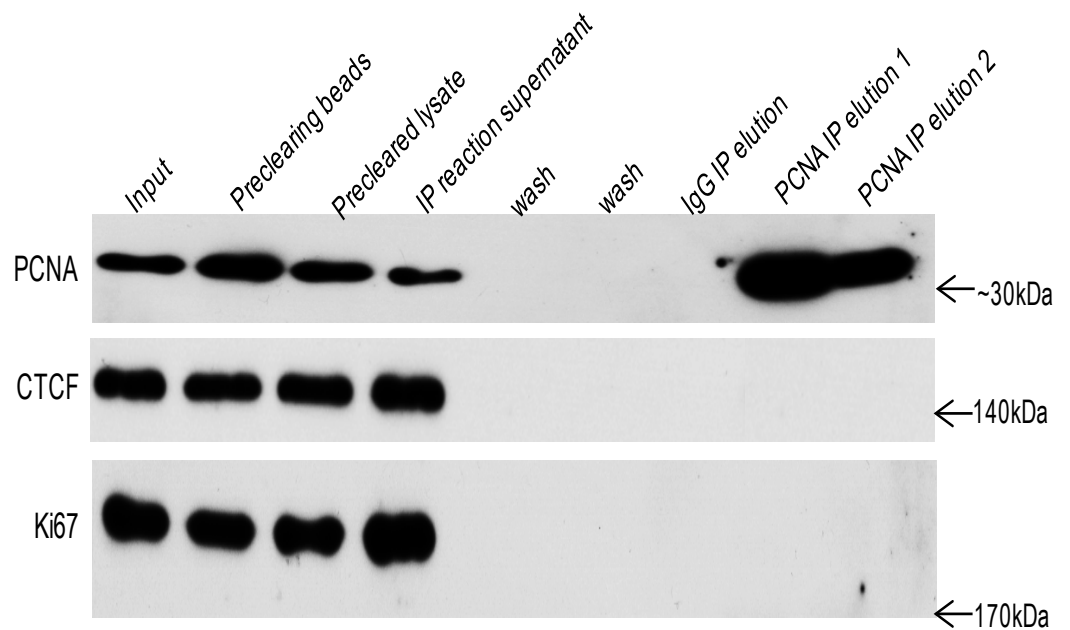
The anti Ki67 antibodies (Vector VP K452 and Abcam 833) did not IP Ki67 protein using the same buffer composition in MCF7 cells (data not shown). On account of the fact that this low stringency buffer was successful in immunoprecipitating CTCF in MCF7 cells and having a lower tendency to disrupt native protein interactions, all other subsequent IPs in this thesis were performed with IP buffer 3.



**Figure 3.12. CTCF IP and co-IP with Ki67 / PCNA and known protein partners RNA pol II / PARP 1, in MCF7 breast cancer cells using a low stringency buffer.** Immunoprecipitated CTCF is shown in the CTCF IP elution lane in (A), but no coimmunoprecipitated protein bands are demonstrated in that lane for Ki67, PCNA and PARP1. A protein band heavier than the band in control lanes is shown in the CTCF elution lane for the RNA pol II (D) panel.

#### **3.4.13. PCNA immunoprecipitation and coprecipitation with CTCF and Ki67 in the MCF7 breast cancer cell line using a low stringency IP lysis buffer.**

Proteins differ in their ability to immunoprecipitate out of a complex lysate (Klenova *et al.*, 2002). Empirical testing is needed to verify the response of a particular protein to the immunoprecipitation process. While CTCF coIP with PCNA was not demonstrated in the previous Figures, it is possible that PCNA might exist in complexes more amenable to co-immunoprecipitation than CTCF. To further investigate whether CTCF and PCNA proteins are physically bound in a complex in the cell nucleus, PCNA protein immunoprecipitation and coprecipitation with CTCF and Ki67 was performed. PCNA protein was immunoprecipitated from MCF7 cell lysates using 20µg of anti PCNA antibody (Abcam 29) and the low stringency IP lysis (buffer 3). An equivalent concentration of IgG antibody was used to IP MCF7 cells and served as negative control. The results shown in figure 3.13 (PCNA panel) revealed that PCNA was successfully IPed as demonstrated in the elution steps. On assessing for coimmunoprecipitation by probing the IP blot membrane with mouse anti CTCF (BD Biosciences; 0.5 µg/ml) and mouse monoclonal antibody to Ki67 (Vector VP-K452, 1:1000 dilution, Ki67), no coimmunoprecipitation with these proteins was demonstrated (figure 3.13, CTCF and Ki67 panels).

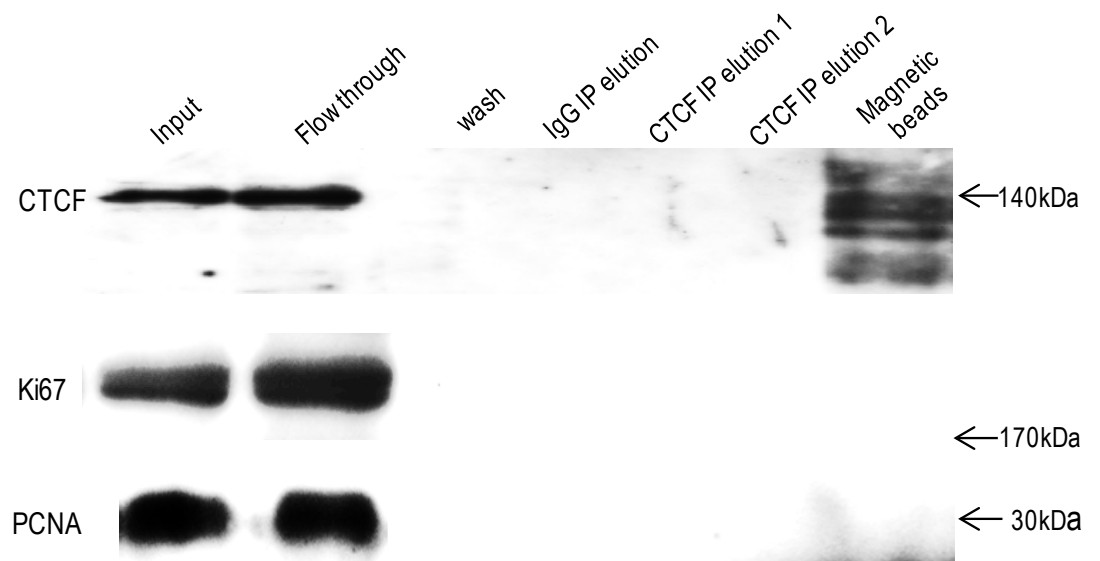


**Figure 3.13. PCNA immunoprecipitation and coprecipitation with CTCF / Ki67 in MCF7 breast cancer cells using a low stringency IP lysis buffer.** PCNA protein was immunoprecipitated in MCF7 cell extracts using 20µg of mouse monoclonal antibody to PCNA (Abcam 29) and a low stringency buffer (Buffer 3). The immunoprecipitated PCNA is shown as a protein band in the 'PCNA IP elution lanes 1 and 2.' There was no PCNA coimmunoprecipitation demonstrated with CTCF and Ki67.



#### **3.4.14 Immunoprecipitation of CTCF and coprecipitation with Ki67 / PCNA using a cross-linker and a magnetic commercial kit**

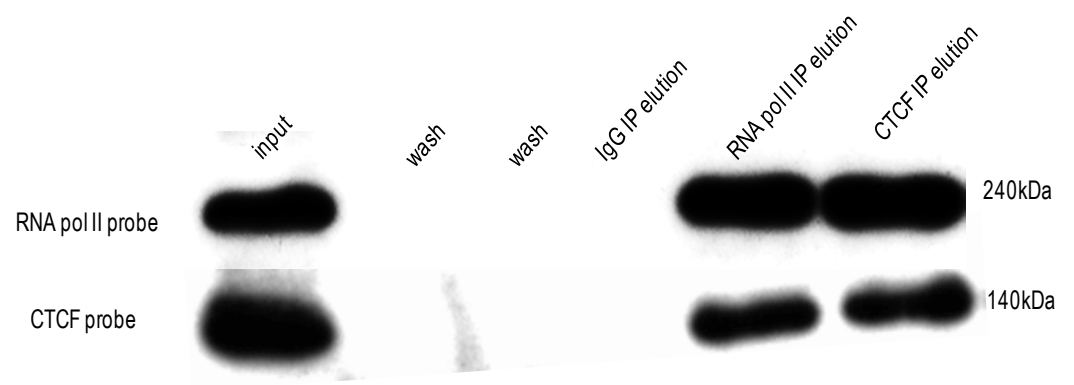
In order to ensure that the lack of CTCF co-IP with Ki67 / PCNA was not due to technical issues (including multiple washes and centrifugation) associated with the traditional immunoprecipitation process, further CTCF IP and coIP were performed using a commercial co-IP kit incorporating magnetic beads and an antibody – bead cross linker (Crosslink magnetic IP / co-IP kit, Pierce). The need for centrifugation was eliminated by the use of a magnet to pull the magnetic beads to the side of the tube. Also the addition of a cross linker eliminated possible antibody fragment interference. CTCF (Millipore) antibody (20 $\mu$ l = 20  $\mu$ g – 60 $\mu$ g) was covalently bound to magnetic beads with disuccinimidyl suberoside (DSS) and the IP process performed first with the IP lysis buffer supplied with the magnetic kit and then later with the low stringency IP lysis (buffer 3) in MCF7 lysates. An equivalent concentration of IgG antibody was used to IP IgG from MCF7 cells and served as negative control experiment. CTCF IP performed with the lysis buffer supplied with the magnetic kit did not detect CTCF in the input and other control lanes (data not shown). On using the low stringency IP lysis (buffer 3) and the rest of the magnetic kit, the input and flowthrough (IP reaction supernatant) showed roughly equivalent levels of CTCF protein as shown in figure 3.14. CTCF was not IPed as there were no protein bands detectable in both elution lanes (figure 3.14). To determine whether protein bands were stuck to the magnetic beads and possibly uneluted, the magnetic beads were boiled in SDS elution buffer and loaded for SDS PAGE. As shown in Figure 3.14, there are protein bands shown in the ‘magnetic beads lane’ corresponding to the molecular weight of CTCF. The elution process performed with other elution buffers was also not successful in demonstrating any possibly immunoprecipitated CTCF protein. In effect, the commercial magnetic kit together with a crosslinker was not superior to the traditional IP process in immunoprecipitating CTCF demonstrating that the traditional IP experimental process was probably not responsible for the lack of CTCF co-IP observed.



**Figure 3.14. CTCF immunoprecipitation (IP) and co-immunoprecipitation (co-IP) in MCF7 cells using a magnetic co-IP kit incorporating a crosslinker.** Shown are protein bands for CTCF in the control lanes (input and flow through) but no immunoprecipitated protein in the elution lanes. The loaded magnetic beads appear to reveal some protein bands in the CTCF panel. This commercial kit could not immunoprecipitate CTCF in MCF7 cell lysates.

#### **3.4.15 RNA pol II immunoprecipitation and coprecipitation with CTCF in the HeLa cervical cancer cell line cell using a low stringency IP lysis buffer (buffer 3)**

The results in figures 3.11C to figure 3.14 have shown a lack of CTCF coIP with both Ki67 / PCNA and also known CTCF partners RNA pol II and PARP1 in the MCF7 breast cancer cell line. The interaction between CTCF and RNA pol II however was discovered in HeLa cells, a cervical cancer cell line (Chernukhin *et al.*, 2006). It is not clear whether the absence of CTCF coIP with RNA pol II in the MCF7 breast cancer cell line is a cell-type specific difference or due to inefficiency of the coIP process. In order to answer this question, HeLa cell lysates, prepared with a low stringency IP lysis buffer (buffer 3) were subjected to immunoprecipitation. RNA pol II was immunoprecipitated with 10µg of RNA pol II (sc 899) antibody from these cells while CTCF was immunoprecipitated with 20µl (20 µg – 60µg) of CTCF (Millipore) antibody. An equivalent concentration of IgG antibody was used to IP immunoglobulin G from the HeLa cells and served as negative control. The blotted membrane was probed serially with rabbit polyclonal anti RNA pol II antibody (sc 899, 0.2µg/ml) and mouse monoclonal antibody to CTCF (BD Biosciences, 0.5µg/ml). As shown in figure 3.15, RNA pol II was successfully IPed from the HeLa cell line extract and co-IPed CTCF. This confirmed a previous finding (Chernukhin *et al.*, 2007) and suggested that the IP set up was efficient and that the RNA pol II / CTCF interaction might be cell type specific, restricted to HeLa cells and not present in the MCF7 breast cancer cell line.

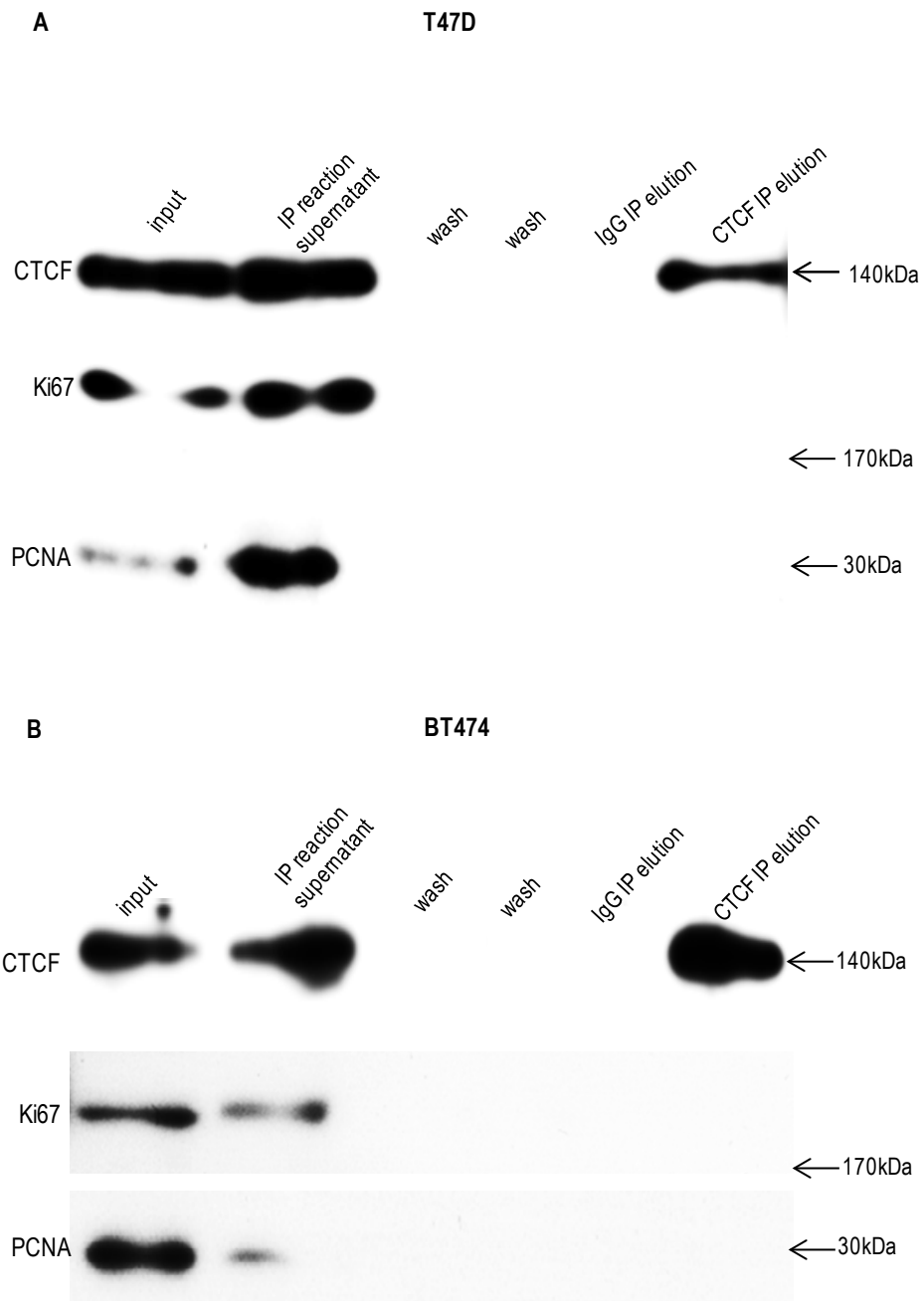


**Figure 3.15. RNA pol II and CTCF coimmunoprecipitation in HeLa cell extracts.**

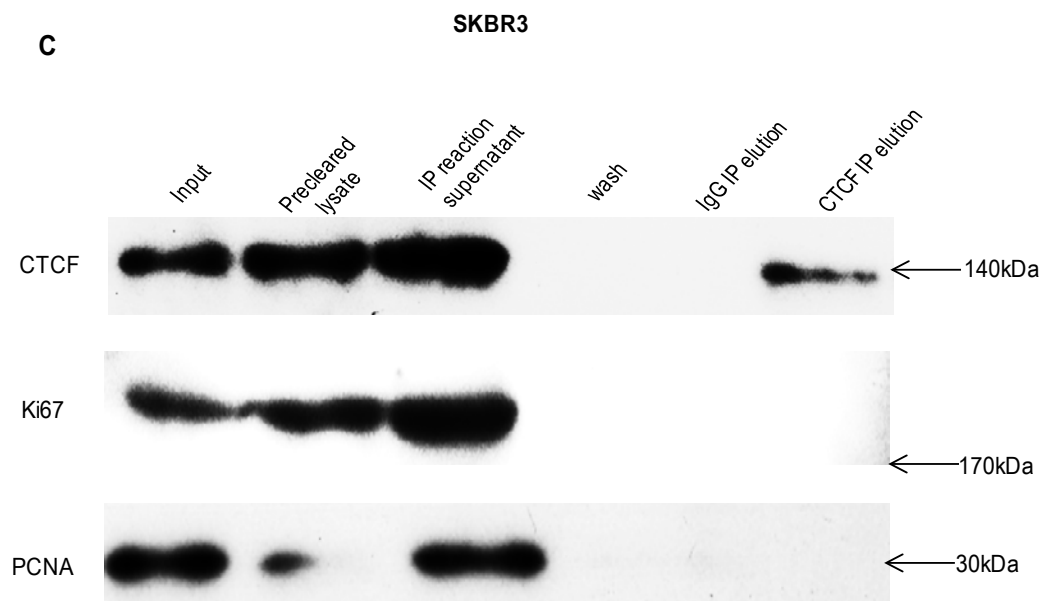
The immunoprecipitated RNA pol II and coprecipitated CTCF bands are shown in the respective elution lanes. The IgG IP elution (negative control) as expected was negative.

#### **3.4.16 Breast cancer cell line - specific differences in CTCF coprecipitation with Ki67 and PCNA in breast cancer cell lines**

Immunofluorescence studies shown in this thesis revealed that CTCF colocalised with Ki67 in all breast cancer cell lines studied and with PCNA in MCF7 and T47D breast cancer cells (figures 3.8 and 3.9) however no physical interaction between these proteins could be demonstrated by co-immunoprecipitation in MCF7 cells. Seeing that CTCF protein interaction could be cell type – specific as shown in the CTCF interaction with RNA pol II in HeLa cells (figure 3.15), and in order to further assess cell type-specific differences in breast cancer cell lines, CTCF immunoprecipitation and co-immunoprecipitation was performed with T47D, BT474, SKBR3 and MDA MB 231 breast cancer cell lysates. CTCF protein was immunoprecipitated from each of the cell lines using 20 $\mu$ l (20  $\mu$ g - 60 $\mu$ g) of CTCF (Millipore) antibody and a low stringency IP lysis buffer (buffer 3). An equivalent concentration of IgG antibody was used to IP immunoglobulin G from the cells and served as negative control. The blotted membrane for each cell line was serially probed with mouse monoclonal antibody to CTCF (BD Biosciences, 0.5 $\mu$ g/ml, A – D top panels), mouse monoclonal antibody to Ki67 (Vector, VP K452, 1:1000 dilution, A – D middle panels) and mouse monoclonal antibody to PCNA (Ab29, 0.3  $\mu$ g/ $\mu$ l, A - D, bottom panels). As shown in figures 3.16 A – D, CTCF was successfully IPed in all cell lines but no co-IP with Ki67 and PCNA was demonstrated. There was also no coprecipitation with RNA pol II and PARP 1 demonstrated (data not shown).



**Figure 3.16. A - D. CTCF IP and co-IP with Ki67 / PCNA in T47D, BT474, SKBR3 and MDA MB 231 breast cancer cell lines.** Figures A to D show immunoprecipitated CTCF band at the expected molecular weight of 140kDa in the CTCF IP elution lane in all cell lines. No coimmunoprecipitated proteins are demonstrated however in the elution lanes of the Ki67 and PCNA panels in all cell lines.



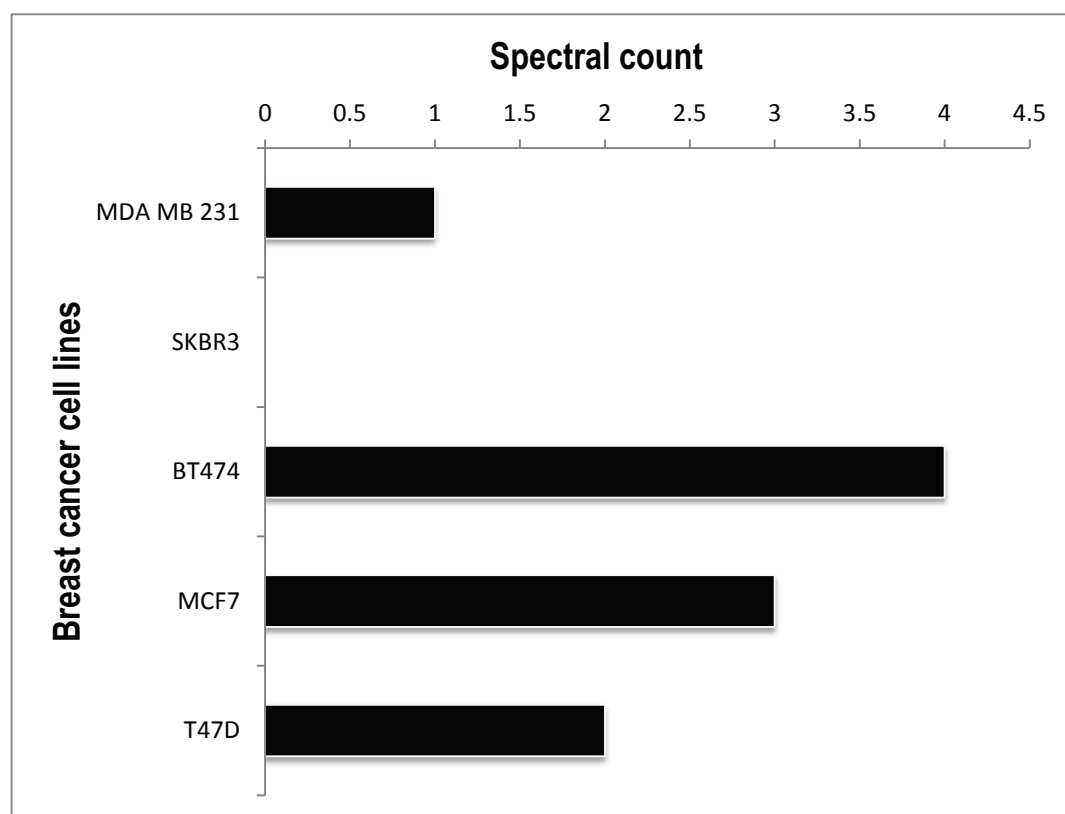
**Figure 3.16 continued**

### **3.4.17 CTCF protein interaction with Ki67 / PCNA assessed by Liquid Chromatography - Mass Spectrometry (LC – MS / MS) in breast cancer cell lines**

Protein interactions need to be tested by at least two independent methods to be valid (Lacroix and Leclercq, 2004). In order to confirm the findings on coimmunoprecipitation and use the most current experimental methods to date for protein interaction assessment, liquid chromatography – mass spectrometry (LC–MS/MS) was performed. CTCF protein was immunoprecipitated with 20µg of Rb mab to CTCF (Millipore, USA) and a low stringency buffer (Buffer 3) in the panel of five breast cancer cell lines. The negative control IgG IP was also subjected to the same process for each cell line. Dr Metodi Metodiev and colleagues performed the mass spectrometry procedure at the regional proteomics centre in Essex University. Briefly, IP lysates for each cell line were first run on a gel and proteins bands were cut out and digested with trypsin. The digests were subjected to liquid chromatography as described in the materials and methods. Open label mass spectrometry was then performed with resultant mass spectra was compared to a protein database. The spectral count indicates the relative abundance of the protein (Lundgren *et al.*, 2010) and as shown in figure 3.17, CTCF was identified in all of the breast cancer cell lines except SKBR3 with spectral count ranging from one for MDA MB 231 to four for BT474. The result supported the bias of CTCF for ER positive breast cancer cell lines considering the spectral count of one in MDA MB 231 and the fact that it was not identified in SKBR3 cells both of which are ER negative breast cancer cell lines. The lack of identifiable CTCF in SKBR3 cells on mass spectrometry while it was abundantly evident on IP and western blotting points to a possible limitation of this mode of investigating protein interactions. This observation provided further evidence for the need to confirm protein interactions with at least two methods. The total number of proteins identified for each cell line and the corresponding spectral counts are shown in table 3.4. Since the mass spectrometric assay was open-label, only proteins demonstrating a spectral count of four or more were regarded as specific CTCF



interacting partners (Lundgren *et al.*, 2010). A detailed inspection of the mass spectrometry data across all cell lines did not identify Ki67 or PCNA as CTCF interacting partners. Interestingly also, neither RNA pol II nor PARP 1 were identified as CTCF - binding partners. Furthermore, no physical binding between CTCF and ER or PR was observed. These findings agree with the results of co-immunoprecipitation assays in figures 3.11 – 3.15 and further confirmed that the interaction between CTCF, RNA pol II and PARP1 discovered in HeLa cells and lymphocytes respectively, was not present in breast cancer cell lines. Vimentin, a marker of breast cancer cell invasiveness (Mendez *et al.*, 2010), was also not identified as an interacting partner of CTCF across all breast cancer cell lines.



**Figure 3.17. CTCF spectral scores on liquid chromatography / mass spectrometry in breast cancer cell lines.** Lysates of breast cancer cell lines after undergoing immunoprecipitation with anti CTCF antibody (Millipore, USA) were subjected to mass spectrometry (see materials and methods, section 2.2.11). The histogram shows spectral counts obtained on label-free mass spectrometry indicating the relative abundance of identified CTCF in each breast cancer cell line.

<b>Breast cancer cell line</b>	<b>Total number of proteins</b>	<b>Number of proteins with spectral count of 4 and over</b>	<b>Number of proteins with spectral count of 3</b>	<b>Number of proteins with spectral count of 2</b>	<b>Number of proteins with spectral count of 1</b>
<b>MCF7</b>	430	-	66	106	258
<b>T47D</b>	492	3	11	28	461
<b>BT474</b>	515	1	3	17	494
<b>MDA MB 231</b>	493	-	34	98	363

**Table 3.4. Distribution of total number of CTCF-interacting proteins identified on label-free mass spectrometry.** Each breast cancer cell line together with the total number of proteins and the corresponding spectral counts are shown. The vast majority of proteins in all cell lines have a spectral count of 1 and were regarded as non-specific. Only those proteins showing a spectral count of 4 and over were considered for further analysis.

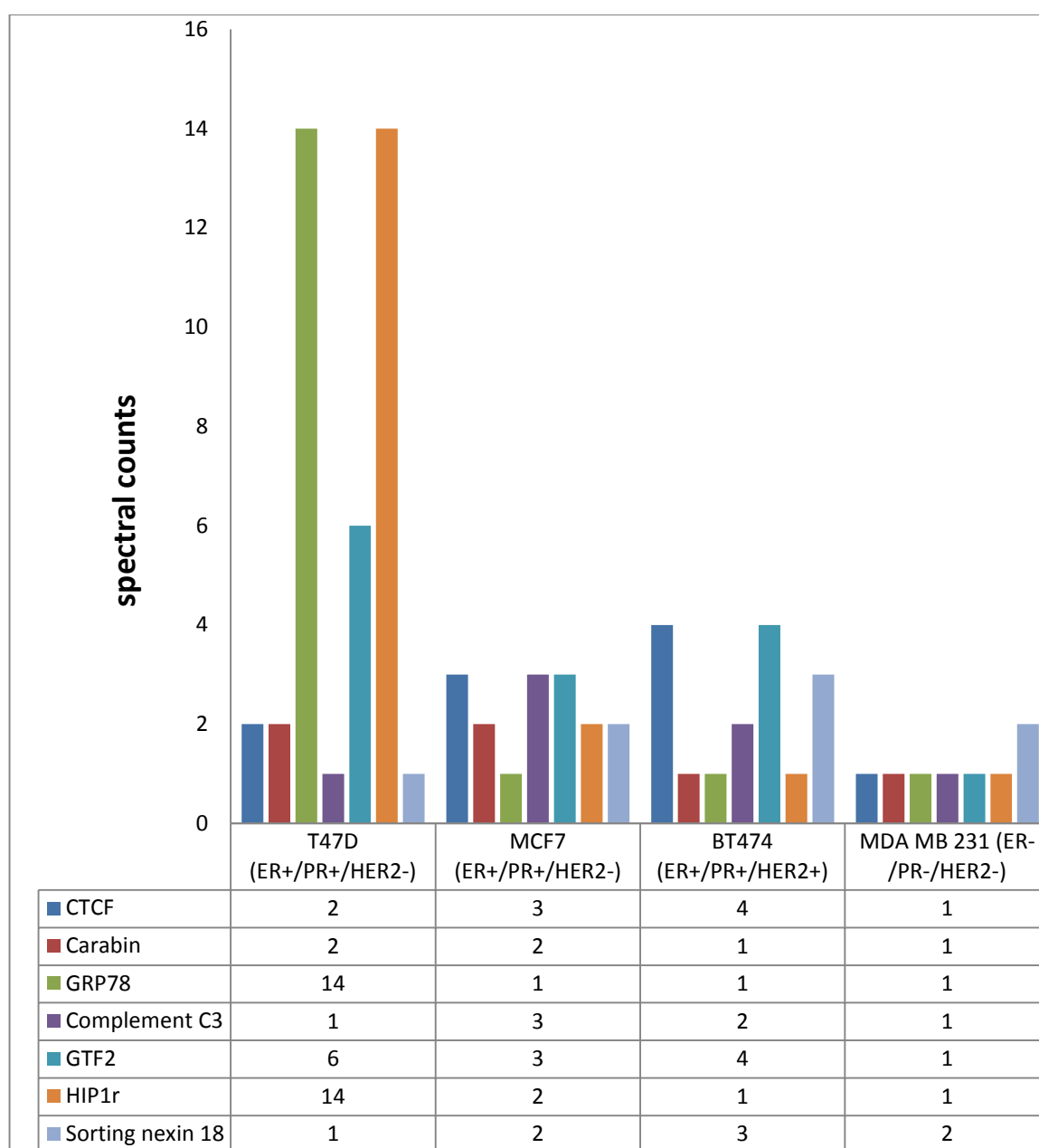
### 3.4.18 Novel CTCF-interacting partners in breast cancer cell lines

For this thesis, a spectral count of 4 and over was regarded as significant interaction (Lundgren *et al.*, 2010). Three proteins satisfied these criteria and are listed in table 3.5. Together with the associated spectral counts, Table 3.5 also shows the main cell line(s) in which the interacting proteins were identified. All three proteins were identified in ER positive cell lines. It is interesting that though the MCF7 and T47D cell lines share the same HR and HER2 biology, none of the three CTCF-interactors identified obtained up to 4 spectral counts in the MCF7 cell line. This emphasised the fact that these two cell lines might represent different disease entities. Looking at the role of these novel CTCF-interacting partners in cancer (table 3.5), there appears to be a unifying connection through epithelial growth factor signalling (EGFR) which is a well-known signalling pathway in cancer biology (Brand *et al.*, 2013)

To further determine whether there was a breast cancer cell - type related difference in CTCF-interacting proteins with respect to abundance across the panel of breast cancer cell lines studied, interacting proteins that were identified in all cell lines are shown in figure 3.18. The appended spectral counts indicate the relative abundance of the interacting proteins in the cell lines. There is a general low abundance of CTCF interacting partners in the triple negative MDA MB 231 cell line with most of the proteins having a spectral count of 1 and thus probably non-specific. The ER positive cell lines (MCF7, T47D and BT474) as a group demonstrated a greater association with the identified proteins possessing spectral counts ranging from one to fourteen. Among the ER positive cell lines, the T47D cell line showed the most numerous spectral counts. The picture as a whole might suggest ER-related differences and a possible bias of CTCF expression for ER positive breast cancer cell lines and lends support to the result in figure 3.17.

	Gene name / Uniprot	MCF7	T47D	BT474	MDA MB 231	Known protein partners	Role in cancer
<b>General transcriptional factor 2 (GTF2)</b>	GTF21_human P78347	3	6	4	1	Interacts with SRF and PHOX1.  Part of BHC histone deacetylase complex	Transcription factor. Regulates PI-3K and TGF $\beta$ signalling (Segura-Puimedon <i>et al.</i> , 2013)
<b>Huntington interacting protein-1 related protein (HIP1r)</b>	HIPR_human	2	14	1	1	Interacts with actin, CLTB and HIP1.	Endocytic protein. Stabilises EGFR and increases EGFR phosphorylation on ligand binding (Ames <i>et al.</i> , 2013)
<b>Glucose regulated protein (GRP) 78</b>	HSPA5	1	14	1	1	Interacts with DNAJC1	Endoplasmic reticulum chaperone. Modulates EGFR signalling (Luo and Lee, 2013)

**Table 3.5 Tabular annotation of novel CTCF interacting partners detected by mass spectrometry.** Also shown is the spectral counts in the breast cancer cell lines where the new partners were discovered, known protein partners of the new proteins and their possible role(s) in cancer formation



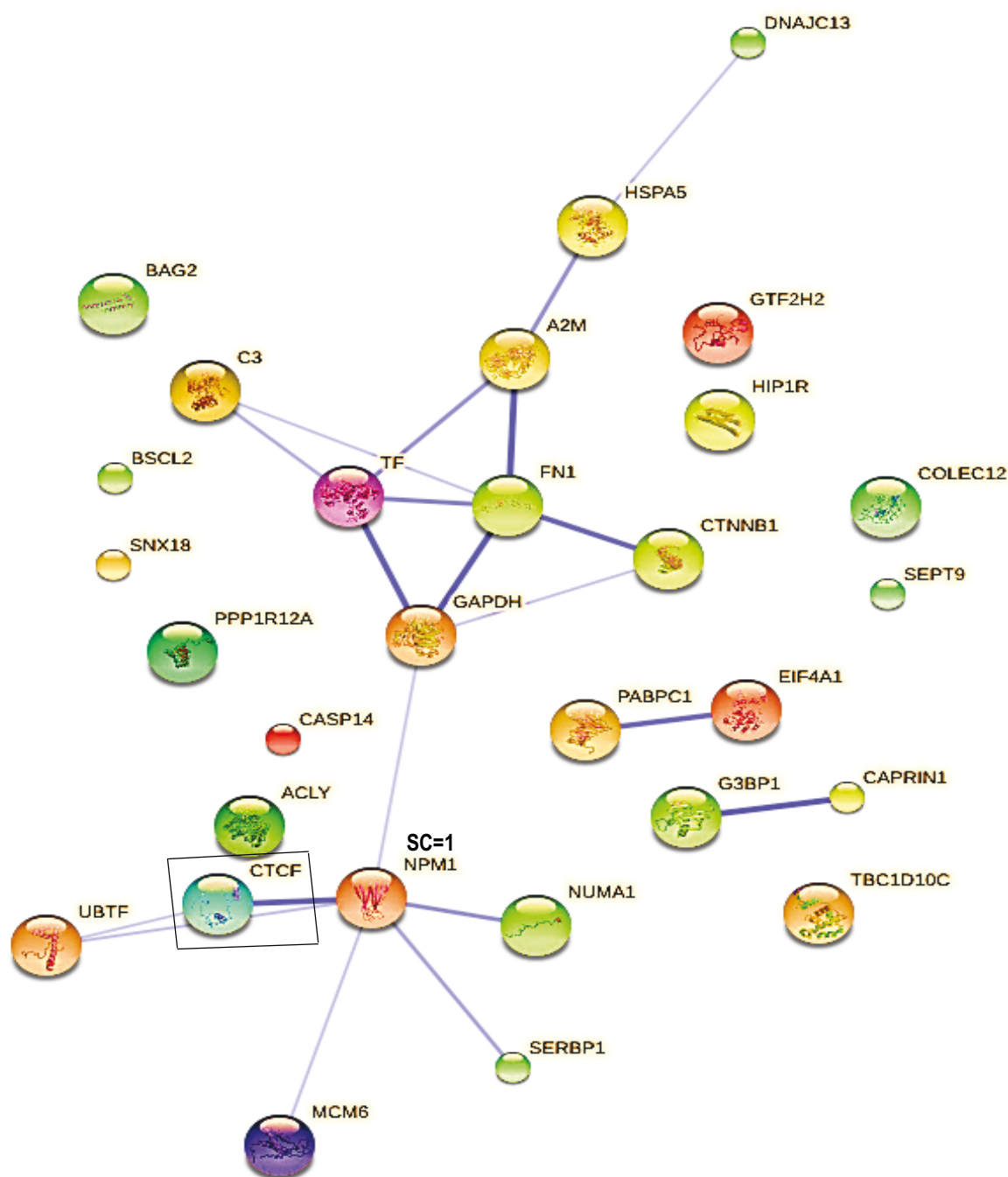
**Figure 3.18. CTCF-interacting proteins identified across all cell lines.** Lysates of breast cancer cell lines after undergoing immunoprecipitation with anti CTCF antibody (Millipore, USA) were subjected to label – free mass spectrometry. The histogram shows CTCF protein interactors identified in all cell lines and their corresponding spectral counts. The spectral counts for the ER positive cell lines for each identified protein compared to the ER negative one is shown. In all cases the negative controls were consistently negative. KEY: GRP78 – glucose related protein 78; GTF2 – General transcriptional factor 2; HIP1r – Huntingtin interacting - protein 1 related.

### 3.4.19 CTCF interacting protein network map in breast cancer cell lines

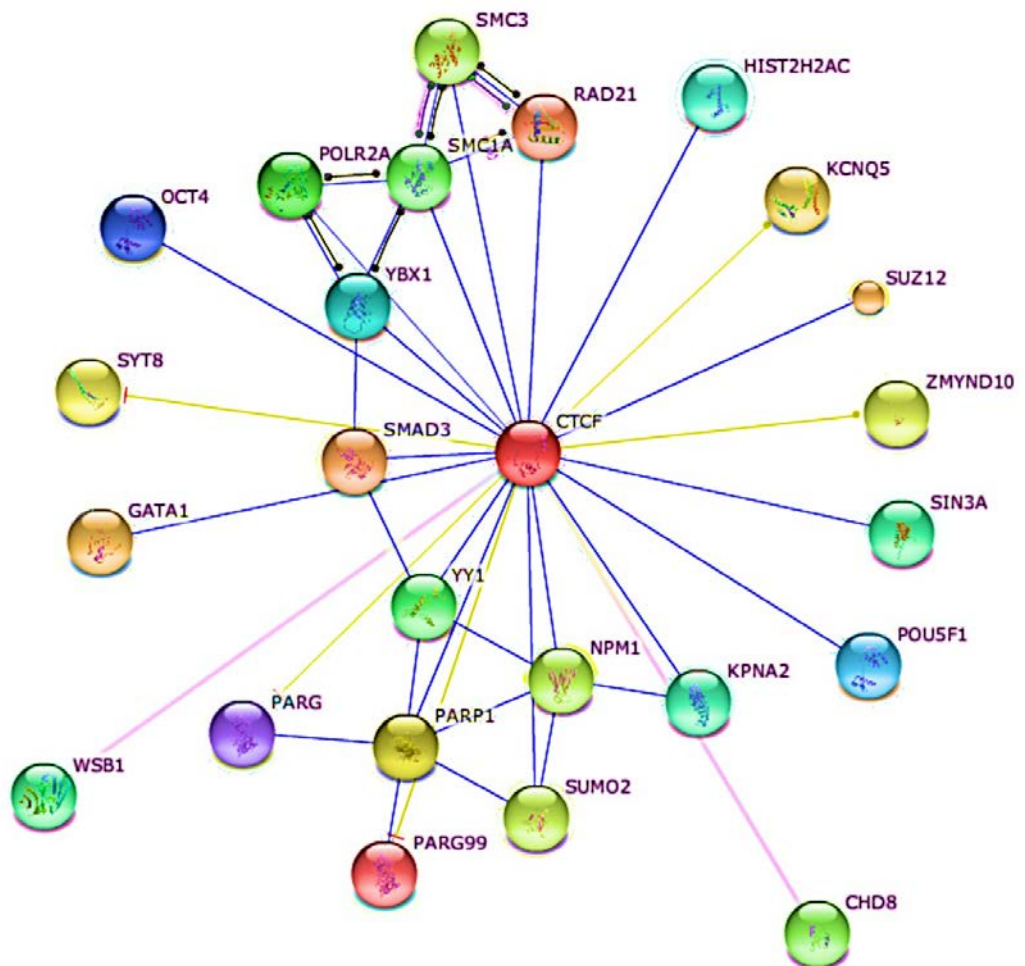
CTCF protein interacting network across human breast cancer cell lines has not previously been described. In order to show this network and how it compared to other CTCF protein interactions, identified breast cancer protein partners were loaded into the Search Tool for the Retrieval of Interacting Genes / Proteins (STRING) software to generate a map (Franceschini *et al.*, 2013). The STRING tool is a database that sources information from previous knowledge in published articles including high throughput experiments and is able to display known and predicted (including physical and functional) protein interactions in over 1100 organisms (Franceschini *et al.*, 2013). Those CTCF interacting proteins with a spectral count of at least two and identified in two cell lines or more were mapped using STRING and the map is shown in figure 3.19. The STRING-generated map of published CTCF protein interactors in *homo sapiens* is shown in figure 3.20. Looking at both figures, the interaction between CTCF and nucleophosmin is shown and the bold connecting line (edge) between both proteins (nodes) (figures 3.19 and 3.20) suggested a functional interaction. The reproduction of this known CTCF - nucleophosmin interaction supported the validity of the results in this thesis. The breast cancer cell line – specific protein network maps generated by STRING are shown in figure 3.21 A - D. The MCF7 map (figure 3.21A) revealed CTCF bound to structural maintenance of chromosome 3 (SMC3), a member of the cohesin complex. This subunit of cohesin is a known CTCF protein interactor in a human T (Jurkat) cell line (Rubio *et al.*, 2008). The bold edge between CTCF and SMC3 suggested that this interaction in the MCF7 cell line could be functional and a further look at the map suggested that this connection might represent the link for other CTCF protein associations in this cell line. This interaction is not seen in the other cell lines and could represent a cell line specific effect. The CTCF – SMC3 interaction further augmented the validity of the results in this thesis. Regarding T47D cells, the edge between CTCF and upstream binding transcription factor (UBTF) indicating physical

binding has not been previously described and seems to be the only link between CTCF and other proteins including GTF2 in figure 3.21 B. The CTCF proteome maps for BT474 and MDA MB 231 cell lines have also not previously been described and are shown in figures 3.21 C and D. Though the maps identify binding interaction between isoforms of GTF2, no edges between CTCF and identified protein interactors are shown demonstrating the lack of published data in this area.

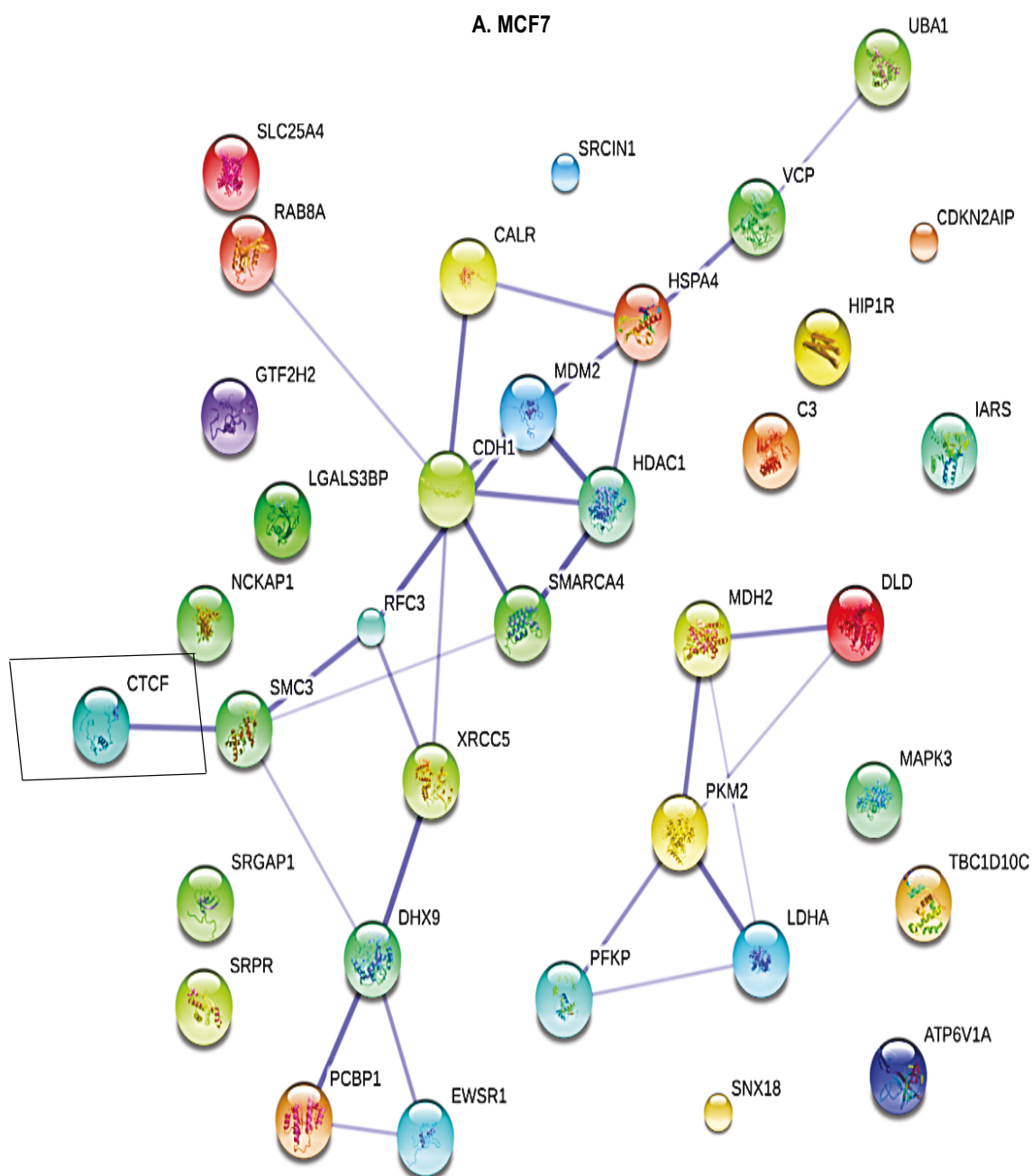




**Figure 3.19. Action view of CTCF interacting proteins identified in at least two breast cancer cell lines with SC of 2 or more.** CTCF is shown in the black box. The bolder the interconnecting line (edge) between proteins (nodes) the greater the probability of a physiological interaction between the nodes. Key: SC – spectral count; NPM1 – nucleophosmin 1; UBTF - upstream binding transcription factor; GTF2 – general transcriptional factor 2. See appendix section 7.1 for full key.



**Figure 3.20. STRING output of known and predicted CTCF interacting protein partners in homo sapiens.** Action view showing edges (predicted functional links) in colour. Nodes represent each labelled protein. Yellow edges depict expression interaction, blue edges depict predicted binding interaction, and pink edges are those interactions that do not fall into a clear category (WSB1 and CHD8). Map generated with a confidence score of at least 0.7. KEY: NPM1: nucleophosmin 1; see appendix section 7.2 for full key.



**Figure 3.21 A – D. Action view of breast cancer cell line - specific CTCF protein interaction network.** MCF7 (A), T47D (B), BT474 (C) and MDA MB 231 (D) breast cancer cell lines protein interactors visualized by STRING. All CTCF interacting partners, in each breast cancer cell line, with a spectral score of 2 and above and negative controls were entered into the STRING software to generate an interacting network map. The colour of the inter-connecting edges is blue representing predicted binding interactions. The bold edges indicate an increased likelihood of a physiological interaction between the proteins. Key: SMC3 - structural maintenance of chromosome 3. Full key in appendix section 7.3 – 7.6.

B. T47D

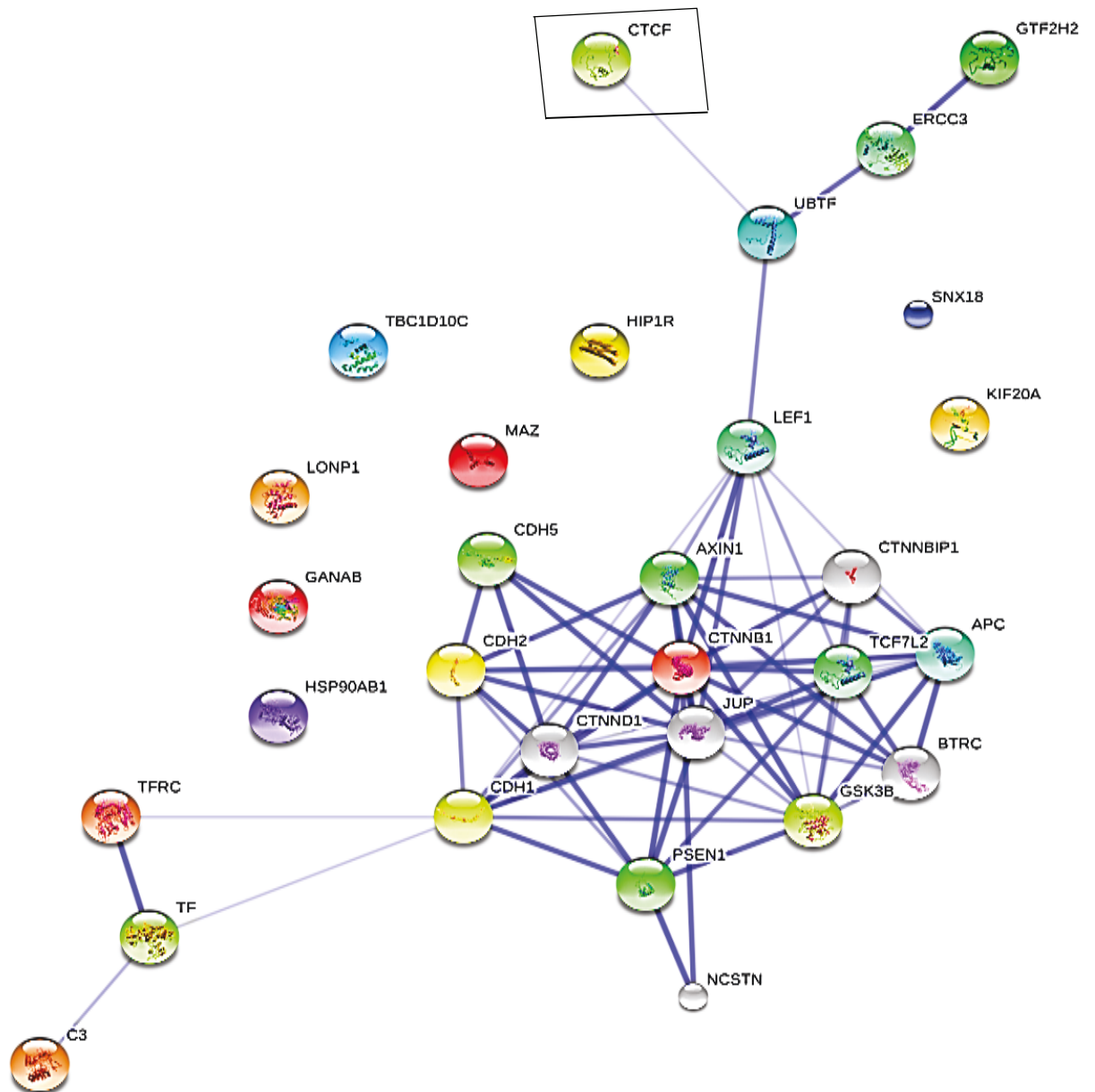


Figure 3.21 continued

C. BT474

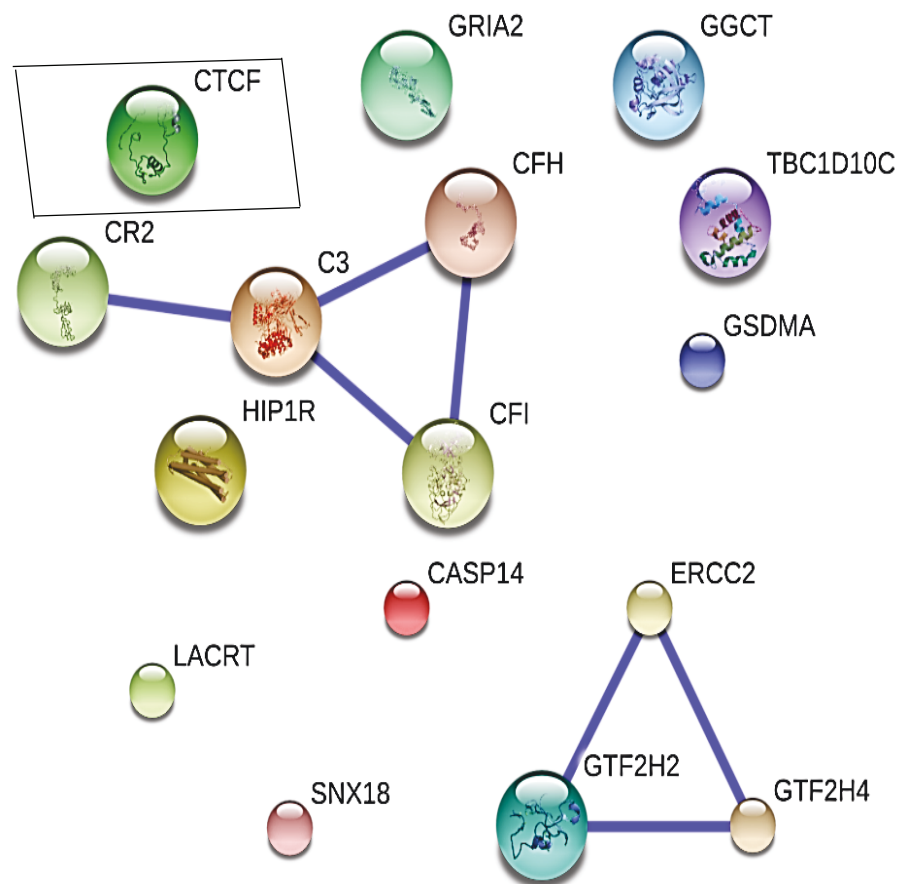


Figure 3.21 continued

D. MDA MB 231

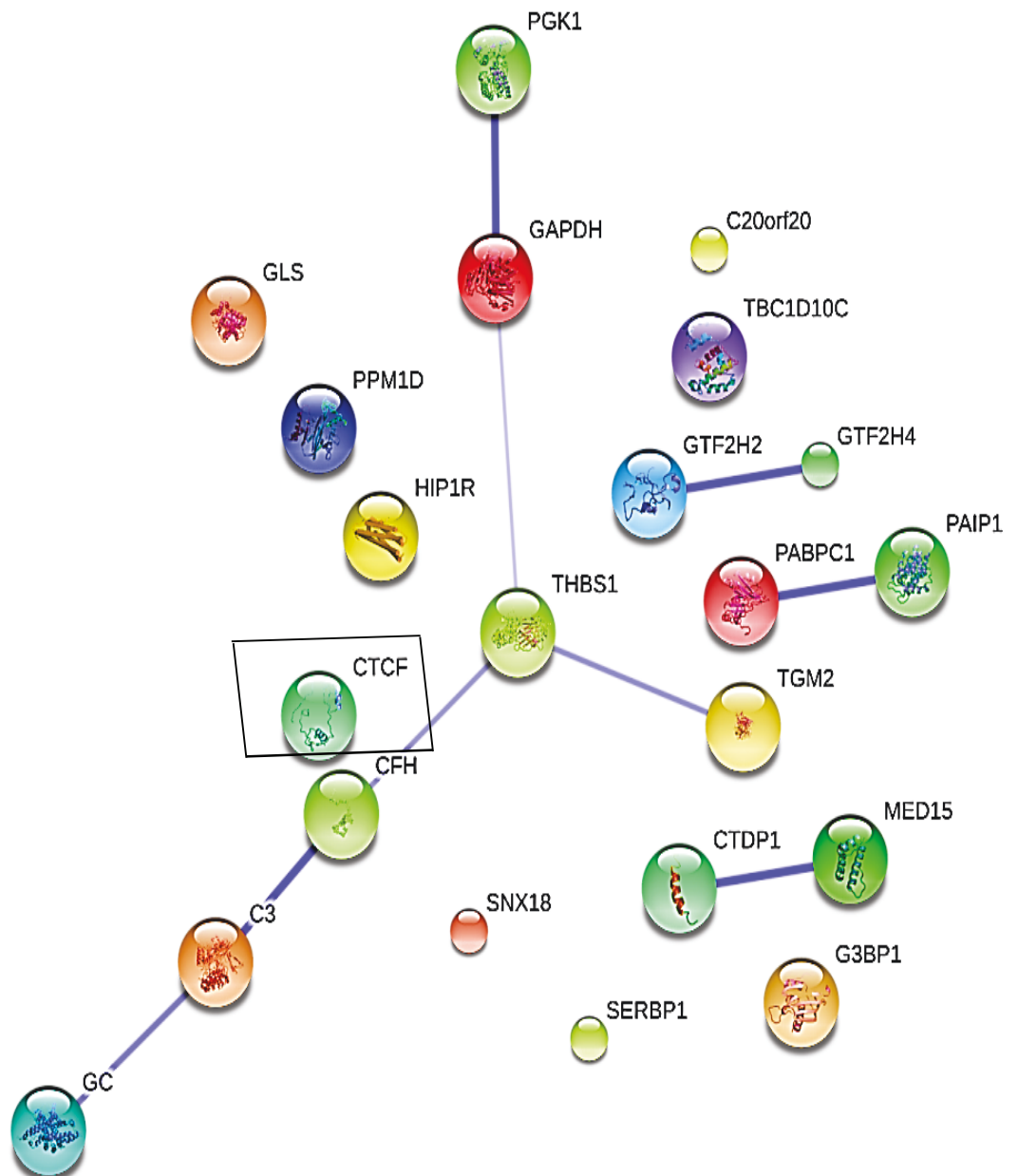


Figure 3.21 continued

## CHAPTER 4

---

### Discussion: Novel CTCF interacting partners

CTCF is a transcriptional regulator with 11 highly conserved zinc finger domains and has been shown to control multiple cellular functions (see section 1.6.3) and undergoes posttranslational modification including poly (ADP-ribosyl) ation (PARylation) (section 1.9.2). There is evidence suggesting that while the poly-PARylated CTCF (CTCF-180) isoform is found in both healthy and cancerous breast tissue, the hypo-PARylated form (CTCF-130) is exclusively present in breast tumours (Docquier *et al.*, 2009). Moreover, a transition from CTCF-180 to CTCF-130 occurs when cells from normal breast tissues are cultured and divide *in vitro*; conversely, CTCF-180 appears following growth arrest in breast cell lines (Docquier *et al.*, 2009). Taken together, this could suggest that the 130kDa CTCF isoform may function as a proliferation marker in breast cancer (Docquier *et al.*, 2009). Other proliferation markers, such as Ki67 and proliferating cell nuclear antigen (PCNA), are valuable as proliferation indicators in breast cancer (Stuart-Harris *et al.*, 2008). Personal communication from F. Docquier suggested that CTCF, Ki67 and PCNA may colocalise in the nucleus of MCF7 breast cancer cells (unpublished data). Moreover, both CTCF and PCNA are known partners of PARP 1 (Frouin *et al.*, 2003; Moldovan *et al.*, 2007; Farrar *et al.*, 2010). It is therefore possible that CTCF and these more established proliferation markers, Ki67 and PCNA, may be linked in breast tumorigenesis. One of the aims of this chapter was therefore to investigate the role of CTCF in breast cancer via its association with known proliferation markers – Ki67 and PCNA. Five breast cancer cell lines exhibiting different hormone receptor / HER2 profiles and with different invasive potentials were studied. The possible interaction between the proteins was studied via indirect immunofluorescence, immunoprecipitation

and co-immunoprecipitation. The findings were then confirmed by liquid chromatography / mass spectrometry.

The results described in Chapter 3 showed that CTCF protein is significantly expressed in the five breast cancer cell lines studied (though less in SKBR3 and MDA MB 231) irrespective of hormone receptor / HER2 status and invasive potential. However multiple CTCF protein bands detected on western blotting with monoclonal and polyclonal anti-CTCF antibodies suggest that different cell lines may express different isoforms of CTCF. Not surprisingly, the isoforms detected appear to be dependent on the antibody as antibodies differ in their ability to detect corresponding proteins due to the variability of their complementarity determining regions (CDRs), which dictates the specificity, and affinity of the antibody for the epitope on the antigen (Fischer, 2014). Using the anti CTCF N-terminal rabbit polyclonal antibody, 160kDa and 130kDa isoforms were detected by western blot across all cell lines with the 160kDa particularly higher in MCF7 cells while SKBR3 cells lacked the 130kDa isoform. The predominant isoform detected with monoclonal anti-CTCF antibodies from BD Biosciences and Millipore had a molecular weight of 140kDa (in agreement with the manufacturers) and was uniformly expressed across all cell lines. The monoclonal anti CTCF antibody from BD Biosciences was also able to detect in addition 100kDa and 110kDa isoforms especially in the hormone receptor positive cell lines (T47D, BT474). The finding of multiple CTCF isoforms in this thesis is in agreement with Zhang *et al.* (2004), who used two-dimensional (2D) gel electrophoresis to study synchronised HeLa cells and identified many isoelectric variants of CTCF. They suggested that their finding could represent a range of PTMs of CTCF protein within different cell cycle stages. Looking at the multiple CTCF isoforms detected in this thesis, it is surprising that no CTCF protein band at 180kDa was identified and is at variance with Docquier *et al.* (2009), who identified this CTCF isoform in MCF7 cells using the same polyclonal anti N Terminal CTCF antibody used in this thesis. It is possible that different batches of the anti N terminal antibody



initially made in chicken and then in rabbit could be responsible for the disparity. Also, the MCF7 cells used for the findings of Docquier *et al.* (2009), may not represent the same cells as cultured cells acquire mutations and change over time (Del Campo *et al.*, 2014). Interestingly, the expressed 160kDa isoform was more abundant than the 130kDa isoform in MCF7 cells and is again in contrast to Docquier *et al.* (2009), who suggested that the predominant isoform in MCF7 cells on western blotting was the 130kDa. Though the same polyclonal antibody to the N terminal region of CTCF was used in both studies, the disparity in the identified CTCF isoforms could be due to the previously mentioned difference in antibody batches, the animal source and changing characteristics of cultured MCF7 cells. Lastly, the significantly lower expression of the 160kDa in the other breast cancer cell lines compared to MCF7 might be due to cell type-specific differences, a factor that might also be responsible for the absence of the 130kDa isoform in the SKBR3 cell line. Taken together, this result showed that there may indeed be multiple CTCF isoforms whose expression could be breast cancer cell line - specific.

To gain insight into the distribution of CTCF in the cell nucleus, immunofluorescence experiments were performed. Biological systems are functionally compartmentalised and the presence of an uncharacterised protein in a compartment of known function generally might give an idea about the function of the uncharacterised protein (Bolte and Cordellieres, 2006). The experiments in this thesis showed that *CTCF detected with the* monoclonal antibody from BD Biosciences (which identified a 140kDa isoform) though expressed in the nucleoplasm, is particularly expressed in the nucleolus of all cell lines studied irrespective of hormone receptor status and invasive potential. Previous work in the K562 leukaemia and MCF7 breast cancer cell lines noted the accumulation of CTCF in the cell nucleolus only when the cells were drug treated (Torrano *et al.*, 2006). This immunofluorescence-based report did not detect nucleolar CTCF in untreated K562 leukaemia cells using anti-CTCF antibody from BD Biosciences (monoclonal) nor in

MCF7 cells using a polyclonal anti-CTCF antibody (Abcam). On western blotting however, they showed that purified nucleolar fractions of untreated MCF7 cells demonstrated CTCF expression suggesting that lack of CTCF detection in untreated MCF7 using IF was not due to the absence of CTCF in the nucleoli of those cells. Results in this thesis showed abundant nucleolar CTCF detected by IF using the same cell line (MCF7) and without any drug treatment. The experimental procedure for IF (including microwaving) in Torrano *et al.* (2006), was the same as in this thesis suggesting that the same antigen epitopes may have been exposed and to the same degree. The disparity between the work of Torrano *et al.* (2006) and this thesis might therefore reside with the monoclonal anti-CTCF antibody from BD Biosciences, which was not used to stain untreated MCF7 cells in the Torrano *et al.* (2006) report. The conclusion that CTCF nucleolar accumulation might be linked to growth arrest (treatment with sodium butyrate) in breast cancer cells as alluded to by Torrano *et al.* (2006), may therefore not be the case if more sensitive and specific anti CTCF antibodies have since been generated. This finding of nucleolar CTCF enrichment could suggest specific nucleolus-associated activity of CTCF in breast cancer cells.

The Ki67 protein, a marker of cellular proliferation, has prognostic significance in breast cancer (Urruticoechea *et al.*, 2005). Most studies of Ki67 protein expression levels have utilised immunohistochemistry (IHC) and employed a cut off of ~14% to define high and low expression levels (Stuart-Harris *et al.*, 2008; Wang *et al.*, 2011). There have not however been internationally agreed standards for methodology including the area(s) of the tumour to sample for microscopy in IHC (Dowsett *et al.*, 2011). If a similar detection rate between immunofluorescence and IHC is assumed, then a 14% cut off for Ki67 expression would mean that all the breast cancer cell lines studied in this thesis possessed a 'high' Ki67 proliferation index as the lowest expression level was 38% (table 3.1). Based on proliferation rates defined by Ki67 expression it would then appear that CTCF expression was not dependent on breast cancer cell line phenotype as all the

cell lines in this study demonstrated high proliferation rates. Adding to the debate is the somewhat surprising marked Ki67 expression in 226LDM cells. This immortalised normal mammary luminal cell line was generated using retroviruses that transduced SV40 large T antigen in addition to the activity of the catalytic subunit of the human telomerase reverse transcriptase (hTERT) enzyme (O'Hare *et al.*, 2001). The high Ki67 expression in this immortalised cell line might be due to the presence of two drivers (virus and telomerase) of proliferation. Further explanation for the difference in Ki67 expression pattern among the five breast cancer cell lines studied might relate to the immunoreactivity of the anti-Ki67 antibodies used. It is recommended that the Molecular Immunology Borstel 1 (MIB1) and Sp6 anti-Ki67 antibodies be used for IHC since they are able to detect a 16-time repeat Ki67-unique epitope motif at the C-terminus (Scholzen and Gerdes, 2000; Dowsett *et al.*, 2011). It is not clear whether this statement is applicable to IF and WB. This thesis however showed that whether the antibody epitopes are identified (Sp6) or not (Ab833 and VP K452 antibodies), Ki67 protein expression is readily identifiable by immunofluorescence and western blotting.

The results in chapter 3 of the thesis sought to determine the relationship between CTCF and proliferation factors, Ki67 / PCNA, in breast cancer cell lines of different immune phenotype and invasive potential. The basic assumption of most studies linking proliferation and invasiveness is that increasing proliferation rate raises mutagenic risk that directly feeds into the inverse relationship between differentiation and proliferation in somatic cells (Zhu and Skoultschi, 2001). In effect, the greater the cycling rate of a tumour the likelier an accumulation of mutations that lead to less differentiated tumour phenotypes which are typically more aggressive. The results presented in this thesis however do not support this hypothesis. Of the five breast cancer cell lines studied, MDA MB 231 is known to be the most aggressive, followed by SKBR3, BT474, and then T47D / MCF7 which are non-invasive based on evidence from migration on Matrigel (Lacroix and Leclercq, 2004). In this thesis, based on proliferation rates determined by

Ki67 expression, the most proliferative cell line was MDA MB 231, followed by BT474, MCF7, SKBR3 and then T47D (table 3.1). In effect the non-invasive MCF7 cell line is equally proliferative as the SKBR3 cell line. This lack of a parallel relationship between proliferation and invasiveness in breast tumours is supported by the findings of Prat *et al.* (2010), who reported that the claudin-low group of breast cancers (for instance, SKBR3) expressed Ki67 more than the HER2 group (for instance, MDA MB 231) though SKBR3 cells are known to be less invasive than MDA MB 231 cells. Though the findings in this thesis suggested that MDA MB 231 cells were more proliferative than SKBR3 cells as expected (table 3.1), MCF7 cells were equally as proliferative as SKBR3 cells, a finding that was not expected.

A possible explanation for this non-parallel relationship between proliferation and invasiveness might relate to the migration-proliferation dichotomy phenomenon (Fedotov *et al.*, 2011). This phenomenon, for which mathematical models have been developed, has it that for some tumour types including gliomas, tumour cell proliferation and migration are mutually exclusive phenotypes. In effect, increased cell cycling rates suppress migration / invasiveness and vice versa. A putative biochemical basis for this interaction has been linked to the epidermal growth factor receptor (Wells, 1999; Brand *et al.*, 2013). It was suggested that the balance between activated cascades of the epithelial growth factor receptor (EGFR) intermediates could determine which direction a cell proceeds at a given time point namely either the erk / MAPK pathway for proliferation (Osaki *et al.*, 2011) or the phospholipase C (PLC $\gamma$ ) pathway for motility (Xie *et al.*, 2010). Attempts have been made to prove or otherwise refute this assumption. Working on the same lung cancer cell line (A549) one study detected cell proliferation on EGF stimulation (Hou *et al.*, 2011), while another study did not (Lauand *et al.*, 2013). The latter report showed that not only did EGF stimulation have no effect on proliferation, but also that there was increased cell motility assayed by wound healing and time lapse analysis. Taken together, these studies seemed to suggest indeed that

whether a cell would proceed to proliferation or invasion might depend on the EGFR cascade of intermediates activated. Though EGFR is overexpressed only in the claudin-low subgroup of breast cancer cells (BCC), its expression in other BC cell lines (higher in the more invasive cell lines) is significant enough to be involved in cell regulation (Zhi *et al.*, 2012). With relation to this thesis, it might be surmised that different breast cancer cell lines, possessing different biologic phenotypes, might observe different proliferation / migration laws. MDA MB 231 and T47D cells could obey the classical positive correlation between proliferation and migration; while MCF7 and SKBR3 might be the prototype exemplifying the proliferation – migration dichotomy. Taken together, Ki67 protein expression, while accepted as a surrogate for proliferation, may not be directly linked to invasive potential in breast cancer cell lines.

The proliferative cell nuclear antigen (PCNA) has clearly defined activities as a sliding clamp at the DNA replication fork enhancing DNA proliferation and mediating DNA repair (Moldovan *et al.*, 2007). The results in this thesis regarding PCNA cellular distribution revealed a diffuse nucleoplasmic expression in all cell lines and also a nucleolar accumulation in the weakly invasive, hormone receptor positive MCF7 and T47D cell lines. Nucleoplasmic and nucleolar localisation of PCNA have been previously described however the functional utility of this protein in the nucleolus is not known and there are no recent reports regarding the cellular distribution of PCNA (Celis and Celis, 1985; Wasseem and Lane, 1990). In this thesis, nucleolar localisation of PCNA appeared to be cell type specific. It could be argued that the absence of nucleolar PCNA in the BT474, SKBR3 and MDA MB 231 cell lines in this study might relate to cell cycle phase. As breast cancer cells in this thesis were not synchronised and would therefore be in different phases of the cell cycle in any spread on a coverslip, the lack of a detectable nucleolar form of PCNA would suggest that cell type-specific differences rather than cell cycle phase was responsible for the differences observed. The cell type - specific difference might in turn be related to estrogen receptor (ER) status. Evidence

was recently presented showing that ER $\alpha$  regulated MCF7 cell proliferation by inhibiting *p21* / *p53* expression while upregulating PCNA and Ki67 protein expression in MCF7 cells (Liao *et al.*, 2014). In effect, the basal state of PCNA expression in ER positive breast cancer cells (BCCs) was thought to be significantly higher than the ER negative ones. As there is a constant flux of proteins between nuclear sub compartments it is theorised that the excess PCNA protein in ER positive cells may simply be non-functionally sequestered in the nucleolus as has been found with the nucleolar accumulation of excess H2B histone which served no clear cut functional purpose in that compartment (Dundr *et al.*, 2000; Leung and Lamond, 2003; Musinova *et al.*, 2011). The lack of nucleolar PCNA in the ER positive BT474 cell line may not be surprising as the HER2 overexpression status in this cell line may diminish to some extent the effect of ER activation (Marchio *et al.*, 2008).

PCNA undergoes posttranslational modifications (PTMs) that can affect its residence in nuclear compartments (Moldovan *et al.*, 2007). Among the post-translational modifications of human PCNA is the EGFR-mediated phosphorylation at tyrosine (Y) 211 which interferes with E3 - PCNA interaction and stabilises PCNA (Wang *et al.*, 2006). The EGFR-dependent control over PCNA stability and the more recently discovered interaction of PCNA with *c-Abl* which is dependent on tyrosine (Y) 211 phosphorylation is thought to be the essential contribution of the PCNA clamp to cell proliferation via DNA synthesis (Zhao *et al.*, 2011). EGFR on the other hand is associated with estrogen-induced MAPK-activation in MCF7 cells and it is possible that ER signalling could be routed through EGFR which amongst other activities aids the phosphorylation of PCNA, a PTM that could affect its intranuclear transport and location (van Agthoven *et al.*, 1994; Zhao *et al.*, 2011). This sequence of events would not be expected in ER negative cells and could explain the pattern of PCNA expression shown in this thesis. With respect to hormone receptor status and colocalisation patterns in breast cancer cell lines, Tan *et al.* (2009), worked with three ER+ breast cell lines

(MCF7, T47D, and ZR75-1), and noted that the estrogen receptor (ER) colocalised with the proliferation factor Ki67 in all of the cell lines. Furthermore, Qiu *et al.* (2005), working with the rat mammary gland discovered that a significant proportion of the cells were ER positive and that they showed a 37% colocalisation with PCNA in the nuclei. There is also evolving evidence showing that CTCF can colocalised with ER and might influence the binding of ER to chromatin (Ross-Innes *et al.*, 2011). Taken together, it is possible that the colocalisation noted in our study between CTCF and both PCNA and Ki67 might be mediated partly through an as yet undefined effect of the estrogen receptor. This might also be part of the reason why colocalisation between CTCF and PCNA was not observed in the ER / PR negative cell lines (SKBR3 and MDA MB 231). Clearly complex mechanisms are operative and there may not be a single unifying mechanism relating CTCF to both Ki67 and PCNA.

Biological systems are functionally compartmentalised and the presence of an uncharacterised protein in a compartment of known function generally gives an idea about the function of the uncharacterised protein (Bolte and Cordellieres, 2006). Co-localisation can be described as complete, complete with different intensities or partial (Bolte and Cordellieres, 2006). Qualitative double immunofluorescence, using a standard wide field fluorescence microscope was used in this study to determine protein colocalisation and has been used extensively in cellular research to assess protein colocalisation (Zhang *et al.*, 2004; Farrar *et al.*, 2005; Torrano *et al.*, 2006; Docquier *et al.*, 2009). The results from this thesis showed a distinct pattern of nucleolar colocalisation of CTCF and Ki67 proteins in all cell lines studied irrespective of hormone status and invasive potential. There was also nucleolar colocalisation of CTCF and PCNA proteins in the hormone receptor positive, weakly invasive cell lines – MCF7 and T47D. Though there was expression of CTCF, Ki67 and PCNA in the nucleoplasmic compartment of all cell lines there was no colocalisation in this region. The nucleolus is known to be active in ribosomal RNA synthesis; cell cycle regulation; storage of nuclear

factors; regulation of tumour suppressor and oncogene activities; and processing of telomerase RNA (Torrano *et al.*, 2006). CTCF is localised to the dense fibrillar and granular regions of the nucleolus (Torrano *et al.*, 2006) while Ki67 is at the nucleolar cortex (Bullwinkel *et al.*, 2006). The exact subnucleolar location of PCNA is unknown. It is possible to speculate that some nucleolus – specific function might link the Ki67 and PCNA proteins to CTCF in breast cancer cells.

While the location of proteins in the same vicinity could suggest a functional link between them a physical connection between those proteins needs to be established to prove that they are part of a functional protein complex. Co-immunoprecipitation (coIP) assays were therefore performed to decide whether the colocalisation of CTCF and Ki67 / PCNA actually meant that they were physically bound. The results in this thesis did not demonstrate a physical interaction between CTCF and Ki67 / PCNA, in the breast cancer cell lines assessed via co-immunoprecipitation. Protein immunoprecipitation (IP) has been extensively used to purify CTCF protein in complex cell lysates (Klenova *et al.*, 2002; Chernukhin *et al.*, 2006). In the IP experiments carried out for this study, a well characterised anti CTCF antibody for IP (Millipore, USA) was used to pull out CTCF from breast cancer cell lysates. In all cell lines, the IPed CTCF was identified on western blots and the control IgG IP was negative. The IP process demonstrated verifiable robustness as the silver stained SDS PAGE gels loaded with CTCF IP samples showed distinct protein bands in the IP elution lanes. Also, the anti PCNA antibody successfully immunoprecipitated PCNA in MCF7 cells while the anti RNA pol II antibody IPed RNA pol II in HeLa cells and demonstrated coIP with CTCF as found in a previous report (Chernukhin *et al.*, 2006). The absence of this same RNA pol II interaction in MCF7 cells suggested cell type-specific differences. Furthermore, it had been demonstrated that CTCF copurified with PARP 1 in coIP assays (Farrar *et al.*, 2010). The results in this thesis have not reproduced that finding likely because a different cell line (B4 cells, generated by a stable transfection of 293T cells with a luciferase-neomycin resistance



construct) was used in the Farrar *et al.* (2010) report and the cells were transfected to overexpress PARP1. It is doubtful that the CTCF / PARP1 interaction in the Farrar *et al.* (2009) report is physiologically relevant as the transfected protein in an elevated and distorted concentration in the cell could possibly mediate interactions that would be absent in more physiological concentrations (Klenova *et al.*, 2002; Trinkle-Mulcahy, 2012). With regards to the findings in this thesis, the lack of coIP between CTCF and Ki67 / PCNA might also be related to the marked difference in their molecular weights. It is thought that protein coIP is more successful when proteins within a smaller range of molecular weights are assessed (Klenova *et al.*, 2002). With CTCF at 140kDa, Ki67 greater than 300kDa and PCNA at 36kDa, the interaction could be missed.

The negative coimmunoprecipitation results could have been due to peculiar issues surrounding IP processes. Liquid chromatography and mass spectrometry (LC – MS/MS) was therefore performed to verify the results of coprecipitation experiments and to identify any novel CTCF binding proteins. The combination of immunoaffinity purification and mass spectrometry is possibly the current gold standard method of choice for detecting and resolving protein interactions (Gingras *et al.*, 2007; Lundgren *et al.*, 2010; Kaake *et al.*, 2010; Paul *et al.*, 2011; Trinkle-Mulcahy, 2012). The method is particularly powerful as one experiment could potentially identify multiprotein complexes and because it is quite sensitive - a common pitfall - the use of appropriate negative controls is very important (Kaake *et al.*, 2010). A mass spectrometer typically comprises an ion source, a mass analyser, a detector and a data system and in this thesis, sample ionisation was achieved with liquid chromatography–coupled ionisation electrospray (ESI) (Alldridge *et al.*, 2008). Peptide ion fragmentation on the other hand was performed with the higher energy C-trap dissociation (HCD) and data generated was used to calculate the molecular weight and relative abundance of peptides in the sample and they were identified against a protein database (Alldridge *et al.*, 2008). Mass spectrometry could be isotope-based or label-free. The latter is relatively easy to

perform and results obtained with it compared favourably with the isotope labelled processes (Lundgren *et al.*, 2010). In this thesis, open label method of LC MS / MS was used and in order to eliminate non-specific interactions, a minimum spectral count of four identified novel CTCF - interacting partners with a high degree of certainty (Hendrickson *et al.*, 2006; Lundgren *et al.*, 2010; Kaake *et al.*, 2010). LC – MS / MS confirmed that neither Ki67 nor PCNA was a CTCF binding partner. Both RNA pol II and PARP 1 were not identified as physically bound to CTCF. These observations confirmed the immunoprecipitation results earlier described. Furthermore, no physical interaction between CTCF and ER or PR was demonstrated.

With a spectral count of at least four and IgG controls that showed no spectra, three novel CTCF-interacting proteins were however identified from mass spectrometry data (Table 3.5). The first of the three is the general transcriptional factor 2 (GTF2), a multifunctional transcription factor which relocates to the nucleus upon growth factor binding and tyrosine phosphorylation, an event critical for its downstream activities (Roy, 2007; Segura-Puimedon *et al.*, 2013). GTF2 physically interacts with Erk / MAPK intermediates and recent evidence suggested that GTF2 regulated target genes in EGFR1, PI-3K / AKT and TGF-beta signalling pathways (Roy, 2007; Segura-Puimedon *et al.*, 2013). These pathways are components of EGFR signalling and are prominent in breast cancer cell proliferation, tumorigenesis, cancer cell invasion and metastasis (Masuda *et al.*, 2012). The second identified CTCF-interacting partner is the huntingtin-interacting protein 1 – related (HIP1r), an endocytic protein, which through stabilising receptor tyrosine kinases and binding to inositides enhances EGFR phosphorylation and alters cellular growth (Hyun *et al.*, 2004). More recent evidence identified a conserved group of four tyrosine residues in HIP1r whose phosphorylation was mediated by EGFR (Ames *et al.*, 2013). The exact contribution of HIP1r to oncogenesis is still unknown. The third identified CTCF interacting partner, is the glucose regulated protein 78 (GRP78), an endoplasmic reticulum chaperone, which was found to be upregulated in breast

cancer where it protected against stress-induced apoptosis, increased angiogenesis and via PI3K / AKT signalling, an EGFR pathway, enhanced cell proliferation and survival (Dong *et al.*, 2008; Luo and Lee, 2013). Further evidence suggested that GRP78 was upregulated by GTF2 in prostate cancer cells (Misra *et al.*, 2009). It is therefore remarkable that CTCF strongly interacted with both proteins in breast cancer cell lines and specifically in the estrogen – receptor positive cell lines as shown in this thesis. CTCF expression is known to increase the resistance of MCF7 cells to apoptosis (Docquier *et al.*, 2005). This function ties in with a similar effect of GRP78 which protected against apoptosis further suggesting a strong link between these proteins (Dong *et al.*, 2008; Luo and Lee, 2013).

To the best of current knowledge this thesis is the first to demonstrate evidence identifying GTF2, HIP1r and GRP78 as novel CTCF - interacting partners in breast cancer cell lines. The discovery of these three new CTCF protein partners represented a significant leap / contribution to knowledge in this area. A very recent publication also discovered the physical association between CTCF and GTF2 in a melanoma cell line (Peña-Hernández *et al.*, 2015). These authors immunopurified CTCF from extracts of the MDA MB 435 melanoma cell line and subjected the lysates to mass spectrometry. They identified GTF2 as a protein partner of CTCF and validated the observation via coprecipitation, reverse immunoprecipitation and colocalisation of the two proteins in this cell line. Furthermore, using coprecipitation they also identified the CTCF–GTF2 interaction in two other cancer cell lines - HCT116 (colorectal cancer) and WEHI (B cell lymphoma) - suggesting that the interaction was not specific to melanoma. Using GTF2 knockdown MDA MB 435 melanoma cells via chromatin immunoprecipitation, Peña-Hernández and colleagues further showed that GTF2 was a key regulator of metabolic processes, an activity that was mediated at least in part by targeting CTCF to promoter regions of genes involved with metabolism. Both proteins also interacted to augment CDK8 recruitment and Pol II phosphorylation on serine 5 (Peña-Hernández *et al.*, 2015)

and it is possible to speculate that sum of these effects on metabolism could serve in some yet to be defined way to maintain cellular transformation. The authors did not however investigate whether this interaction also occurred in noncancerous normal cells of the skin, large bowel or lymph glands to determine whether it was a cancer-specific effect. The paper of Peña-Hernández *et al.* (2015) clearly validated the CTCF-GTF2 interaction in three different cancer cell lines and would provide proof that the discovery of the CTCF-GTF2 interaction in breast cancer cell lines as shown in this thesis was real. Since CTCF-140 protein expression in normal breast cells is essentially non-existent as shown in LDM226 normal breast cells in figure 3.2, this interaction with GTF2 is likely to be specific for breast cellular transformation.

The discovery of the CTCF-GTF2 interaction in ER positive (and not ER negative) breast cancer cell lines in this thesis is intriguing. The very low spectral count achieved for CTCF in MDA MB 231 cells might suggest that CTCF was of low abundance in this cell line. It is not clear whether this would explain the lack of copurification with the three new protein partners in this cell line. The ER status of the melanoma, colorectal cancer and lymphoma cell lines used in the report of Peña-Hernández *et al.* (2015) is not known. It is therefore not possible to determine the influence or not of estrogen receptor status on the CTCF-GTF2 interaction in the cell lines used by Peña-Hernández *et al.* (2015). While acknowledging the import of the findings in this thesis and discovering that other authors had independently discovered and validated the physical association between CTCF and one of the protein partners (GTF2) discovered in BC cell lines in this thesis, the thrust of this thesis however revolved around the collective evidence suggested by the three new CTCF protein partners identified (GTF2, HIP1r and GRP78) regarding how CTCF exerted its effects in breast cancer. As with GTF2, further work would involve validating the relationship between CTCF and both HIP1r and GRP78.

A careful inspection of the mass spectrometry results revealed a clear bias of CTCF expression and protein interaction for ER-positive breast cancer cell lines and this

supported the report of Ross-Innes *et al.* (2011). Furthermore, since published reports showed that the novel CTCF interacting proteins exhibited an extensive and specific involvement with EGFR signalling as their predominant mode of action in oncogenesis (Table 3.5) it could therefore be surmised that the mechanistic involvement of CTCF in breast tumorigenesis could be mediated at least in part via EGFR signalling. As noted previously, the three CTCF- interacting proteins were identified in estrogen receptor (ER) positive breast cancer cell lines and the literature indicated that there was an extensive interaction between estrogen and EGFR signalling. For example, estrogen via ER $\alpha$  regulated cellular migration through direct effects on EGFR signalling (Xie *et al.*, 2010). Though there appeared to be an inverse relationship between EGFR and ER $\alpha$  expression in breast cancer cells, both ER $\alpha$  and EGFR remained functional whether one was dominantly expressed or not (van Agthoven *et al.*, 1994; Zhi *et al.*, 2012; Tsonis *et al.*, 2013). Further evidence suggested that uterine DNA synthesis mediated by EGF was abolished in ER $\alpha$  – knockout mice even with wild type levels of EGF and EGFR (Skandalis *et al.*, 2013). Taken together, it would appear that ER $\alpha$  was needed in some EGF-mediated activities and might further suggest that ER $\alpha$  was some kind of centre around which some EGF / EGFR mediated activities revolved. It is therefore not exactly clear whether the three CTCF protein interacting partners discovered in this thesis point to a direct CTCF - EGFR interaction or an indirect interaction involving ER $\alpha$ . It would appear that interaction in ER positive breast cancer cell lines was a common denominator among the new protein partners found in this thesis and there was no direct physical binding between CTCF and ER demonstrated with the mass spectrometry results in this thesis. A possible direct regulatory effect by CTCF on ER $\alpha$  expression has not previously been described.

## CHAPTER 5

---

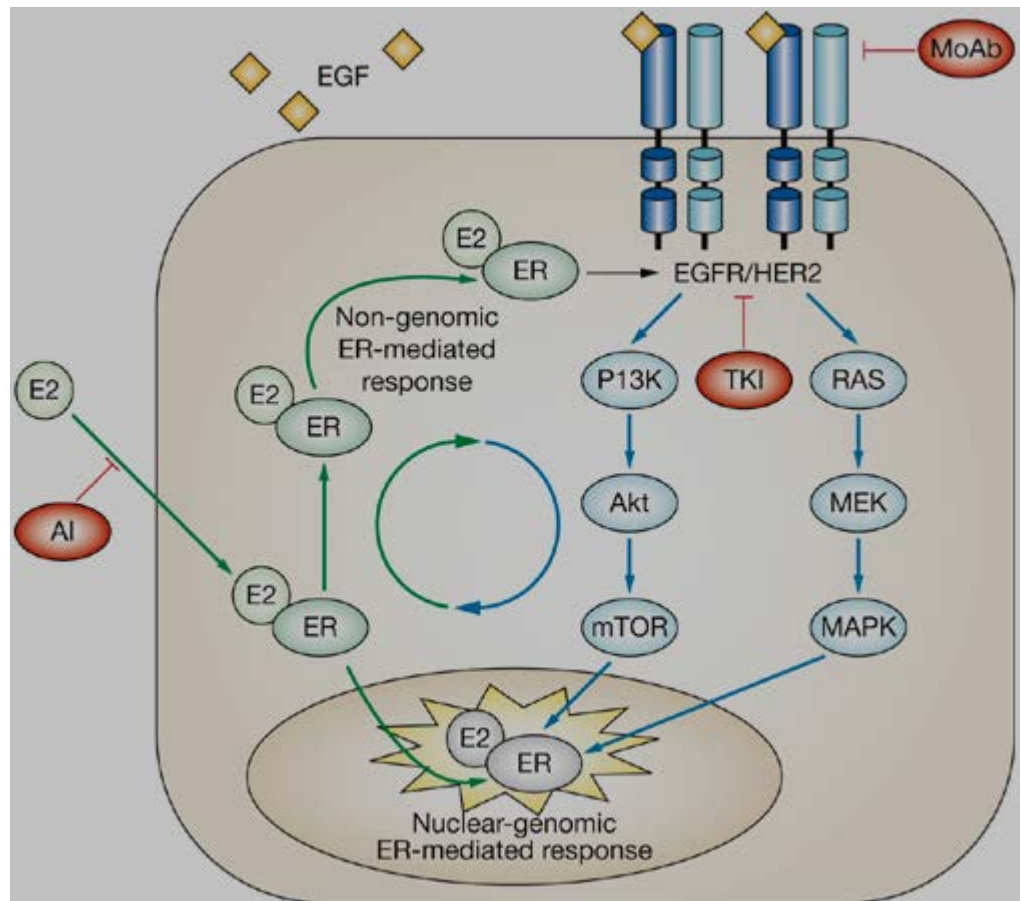
### Regulatory relationship between CTCF and ER $\alpha$

#### 5 Background

The previous sections in this thesis explored the mechanisms through which CTCF was involved in the breast cancer phenotype via investigating a protein interaction with other known proliferation factors in breast cancer. Immunofluorescence, immunoprecipitation, western blotting and open label mass spectrometry were used to investigate this association. The results presented in chapter 3 and discussed in chapter 4 showed that though CTCF colocalised with Ki67 in all cell lines and with PCNA in the ER positive weakly invasive cell lines, there was no physical binding demonstrated between these proteins using immunoprecipitation assays and confirmed by mass spectrometry. New CTCF-interacting proteins were however identified with a high degree of certainty and were seen only in the estrogen receptor positive breast cancer cell lines. Interrogation of these novel protein partners revealed that they interfered with EGFR signalling cascades in cancer formation. As shown in figure 5.1 there is a close relationship between EGFR signalling and the estrogen receptor and as pointed out in section 1.12, CTCF has extensive interactions with ER. It is therefore not clear whether the interaction of CTCF with EGFR (via novel protein partners) is direct or an indirect activity through the estrogen receptor. This section of the thesis addresses whether there was a direct regulatory effect of CTCF on ER expression.

#### 5.1 Role of estrogen in breast cancer initiation

Breast cancer develops more frequently in situations of prolonged estrogen stimulation as occurs with early menarche and late menopause (Barrett, 2010). A direct causal link between estrogen exposure and breast cancer was demonstrated by the transformation



**Figure 5.1. Inter-relation of estrogen, ER and EGFR receptor signalling cascades.** Demonstrated is the link between estrogen and EGFR signalling cascades in cellular cytosol and nucleus. Also shown are the points of inhibition by MoAb, TKI and AI. Key: E2 - estrogen; MoAb- monoclonal antibodies; AI – aromatase inhibitors; TKI – tyrosine kinase inhibitors; PI3K – phosphatidylinositol 3 kinase. Source: Prat and Baselga, 2008.

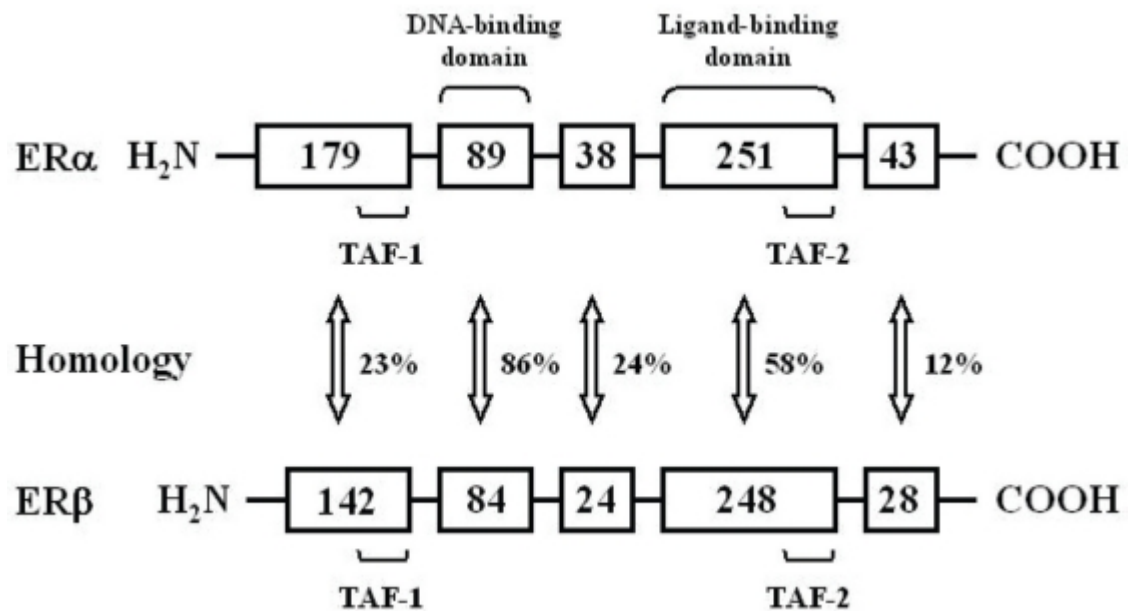
of MCF-10F cells which are immortalised estrogen – and progesterone - receptor negative, normal breast epithelial cells (Russo and Russo, 2006). In this paper, MCF-10F cells were transformed on exposure to estrogen as evidenced by the loss of their usual duct pattern, acquisition of anchorage-independent growth and by becoming invasive on Matrigel. The authors also noted that the transformed cells induced palpable tumours when injected into severe combined immunodeficient (SCID) mice. The report gave no suggestions on how E2 gained entry into the estrogen receptor (ER) negative MCF-10F cells but attributed estrogen-induced neoplastic transformation to direct genotoxic effects. This study supported the results of previous work performed on MCF-10A cells, another immortalised estrogen – and progesterone - receptor negative, breast epithelial cell line which were also transformed on estrogen stimulation (Liu and Lin, 2004). A more recent report seeking to define further the impact of estrogen and its metabolites on breast cancer risk assessed 277 postmenopausal patients in a prospective case – control study (Fuhrman *et al.*, 2012). Using liquid chromatography – tandem mass spectrometry, they measured serum levels of parent estrogens (estradiol, estrone) and their metabolites, and confirmed a statistically significant association between unconjugated estradiol and breast cancer risk. Furthermore, they showed that high 2-hydroxylation of parent estrogen was associated with lower breast cancer risk while the greater the methylation levels in the 4-hydroxylation pathway, the higher the risk for breast cancer. This study corroborated the cell line work previously mentioned (Liu and Lin, 2004; Russo and Russo, 2006) and though not generalizable seeing that premenopausal women were not included, it served to augment the body of knowledge directly linking estrogen to breast cancer risk. Fuhrman *et al.* (2012) also supported the previously held view that estrogen had both a direct mitogenic effect whereby it augmented proliferation of breast cells and also served as a precursor to a potent mutagen – namely excessively methylated 4-hydroxylation products (Yager and Davidson, 2006; Fuhrman *et al.*, 2012). While these reports suggested a clear role for



estrogen in breast cancer initiation, they do not give any answers as to how estrogen gained entry into ER negative cells.

## **5.2 ER structure and activation**

The effects of estrogen on breast cancer development are mediated via its actions at the estrogen receptor on the cell membrane (Acconcia and Marino, 2011). The estrogen receptor is a transcriptional factor that has two main subtypes – ER $\alpha$  and ER  $\beta$  – which have 96% amino acid homology at their DNA binding domain (figure 5.1) but only 53% amino acid similarity in their ligand binding regions (Yager and Davidson, 2006; Nasu *et al.*, 2008). The difference in the ligand binding domain might be responsible for the difference in ER $\beta$  activity which includes an inhibition of both ER $\alpha$ -mediated transcription and estrogen-induced proliferation in breast cancer cells (Strom *et al.*, 2004). Aside from the DNA (C) binding domain which mediates receptor dimerization, ER $\alpha$  also possesses other domains including the N terminal domain which is involved in protein-protein interactions; D domain, the heat shock protein (HSP) binding site; C-terminal domain which binds estrogen and is involved in gene transcription; and the F domain which is involved in ER transcriptional activity (reviewed in Ascenzi *et al.*, 2006). The classification of breast cancer had been based primarily on nuclear ER status determined by immunohistochemistry (Park *et al.*, 2012). As previously mentioned elsewhere, this classification served to identify tumour biology and clinical behaviour and thus stratified patients into specific treatment and surveillance groups. Recent evidence indicated however that the estrogen receptor is not only nucleus-bound but also membrane-bound and cytoplasmic (Ford *et al.*, 2011; Acconcia and Marino, 2011). This could explain why traditionally ER negative breast cancer cell lines have been found to be ER positive on flow cytometry and western blotting since whole cell lysates were tested with these methods (Ford *et al.*, 2011). On estrogen ligand binding to ER $\alpha$  there



**Figure 5.2. Structure and comparison of estrogen  $\alpha$  and  $\beta$  receptor domains.**

The estrogen receptor consists of six functional domains. The numbers in the boxes indicate numbers of amino acids. Homology between the distinct domains of the receptors is noted as percentages. TAF: transcription activating function.

Source: Nasu *et al.*, 2008.

are distinct immediate (extranuclear) and late (nuclear bound) effects. The extranuclear effects are observed within seconds to minutes of binding and include the activation of signalling pathways shown in figure 5.2. These intermediate signals are known to regulate proliferation, survival and apoptosis in transformed cells. The nuclear effects, noticed after about two hours of estrogen ligand binding, result from receptor translocation to the nucleus where it binds DNA at the estrogen response element (ERE) and induces gene transcription (Acconcia and Marino, 2011). The observed consequence of estrogen binding is the sum total of both membrane and nuclear bound ER effects and alteration in rapid ER-induced signalling and / or delayed ER-induced nuclear transcription could lead to dysregulation in cellular control mechanisms and subsequent malignant transformation. Seeing that ER $\alpha$  is involved in many cellular processes, its activity undergoes cellular regulation.

### **5.3 ER $\alpha$ regulation**

The activity of ER $\alpha$  is itself regulated by multiple mechanisms including posttranslational modifications (PTMs) like palmitoylation and phosphorylation, activity of coactivators and corepressors, and epigenetic mechanisms like methylation and histone modifications (Marino and Ascenzi, 2008; Anbalagan *et al.*, 2011). Palmitoylation, by localizing ER $\alpha$  to the cell membrane plays an important role in ER-associated downstream physiological processes (Marino and Ascenzi, 2008). Phosphorylation occurs at multiple sites on the ER $\alpha$  including serine, threonine and lysine residues (Murphy *et al.*, 2011). Phosphorylation is associated with multiple roles including transcription, nuclear localization, and co-activator recruitment (Anbalagan *et al.*, 2011; Murphy *et al.*, 2011). The association of ER $\alpha$  with coactivators and corepressors also partly regulates its own activity. Through their leucine – rich motifs, coactivators bind and induce ER-induced signalling via chromatin remodelling, RNA pol II induction or by involving the basal transcriptional machinery (O'Malley and Kumar, 2009); while corepressors such as histone deacetylase (HDAC), prevent this activity (Kawai *et al.*, 2003; O'Malley and

Kumar, 2009). In some cases of breast cancer, ER is not expressed and there are suggestions that epigenetic processes including hypermethylation and posttranslational histone modification (PTM) (methylation, phosphorylation or acetylation of histone lysine residues) could be responsible for the lack of ER expression (Hervouet *et al.*, 2013). Ramos *et al.* (2010) noted that 41% of sporadic breast cancer samples demonstrated *ER* gene promoter methylation. They confirmed that increasing gene methylation was associated with progressive decline of the *ER* $\alpha$  gene product. Histone modification including lysine acetylation has also been shown to regulate ER expression in breast cancer cells. Overexpression of the histone deacetylase 1 (HDAC1) abolished *ESR1* expression in MCF7 cells (Kawai *et al.*, 2003; O'Malley and Kumar, 2009) and conversely, HDAC inhibitors restored both *ER* $\alpha$  expression and sensitivity of ER negative breast cancer cells to the aromatase inhibitor, letrozole (Sabnis *et al.*, 2011).

## 5.4 Knowledge gap and hypothesis

There is a body of evidence linking ER $\alpha$  expression to CTCF. It is known that ER $\alpha$  regulates MCF7 cell proliferation by inhibiting p21 / p53 expression while CTCF binding has been demonstrated at the promoter of the *TP53* gene where it stops the spread of repressive histone PTMs (Soto-Reyes and Recillas-Targa, 2010; Liao *et al.*, 2013). Also, as mentioned previously, CTCF is known to negatively regulate the forkhead protein (FOXA1 / HNF3 $\alpha$ ), which is required by ER $\alpha$  for chromatin binding (Hurtado *et al.*, 2011). Furthermore, CTCF and ER binding regions colocalise and both proteins recruit HDAC to repress gene expression (Lutz *et al.*, 2000; Kawai *et al.*, 2003). Finally, there is evidence suggesting a bias of CTCF interaction for ER positive breast cancer cell lines (Ross-Ines *et al.*, 2011) an observation supported by the findings of this thesis in the results described in chapter 3. There is however no information directly relating CTCF to ER $\alpha$  and this chapter therefore hypothesized that CTCF may directly regulate *ER* gene expression in ER-positive breast cancer cells.

## 5.5 Objectives of this chapter

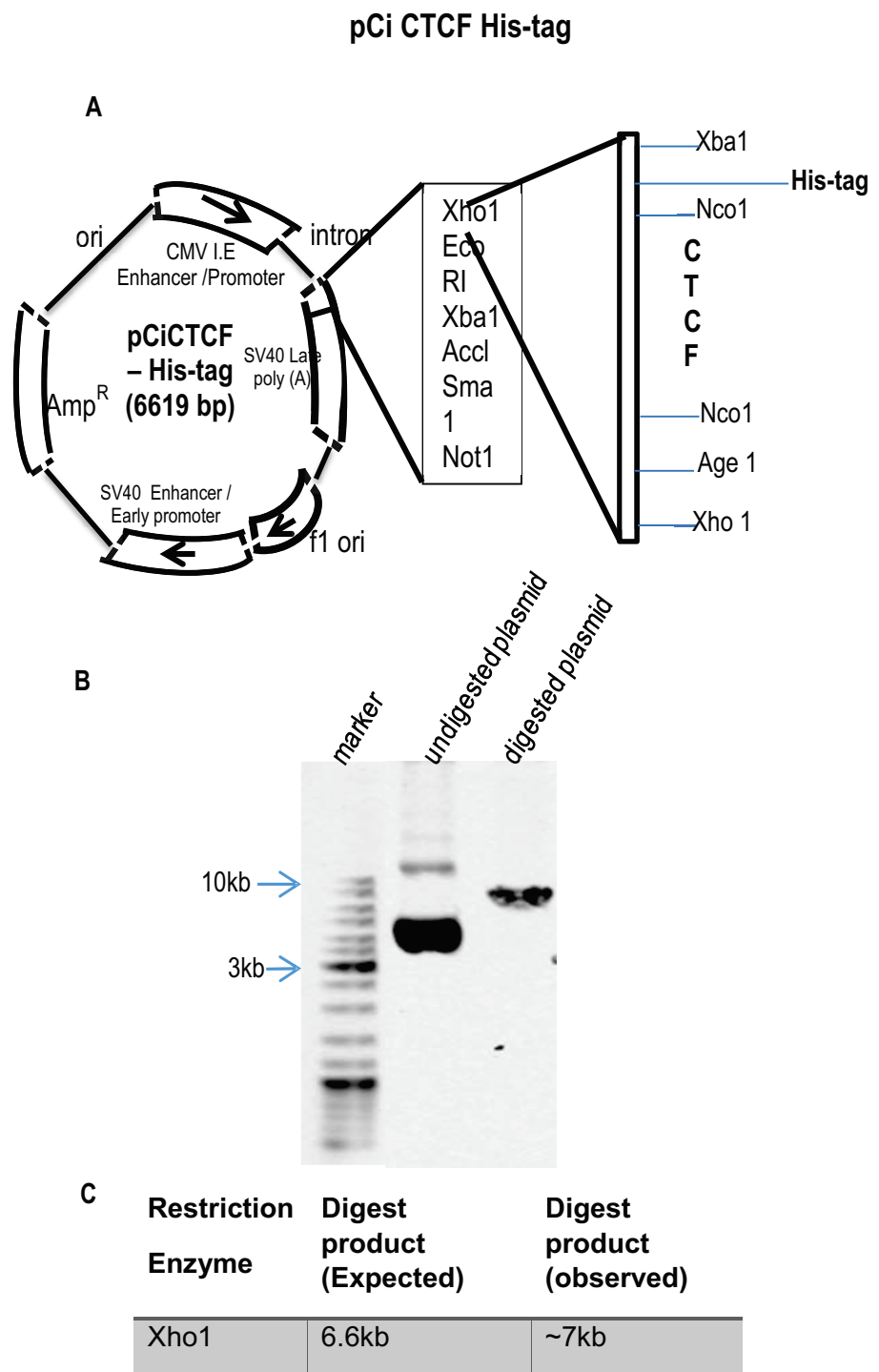
The main objective of this chapter was to determine whether a direct regulatory relationship existed between CTCF and the estrogen receptor (ER)  $\alpha$ . In order to establish whether the expression of ER $\alpha$  product in breast cancer cells was dependent on CTCF, ER $^{+}$  / PR $^{+}$  MCF7 breast cancer cells were transfected with either *CTCF* expression vectors or siRNA against CTCF. Following CTCF over-expression and knockdown, changes in endogenous expression of *ER* and *PR* gene and protein expression were monitored by quantitative polymerase chain reaction (QPCR) and western blot analysis.

## 5.6 Results

### 5.6.1 Restriction enzyme digest of CTCF plasmid expression vectors

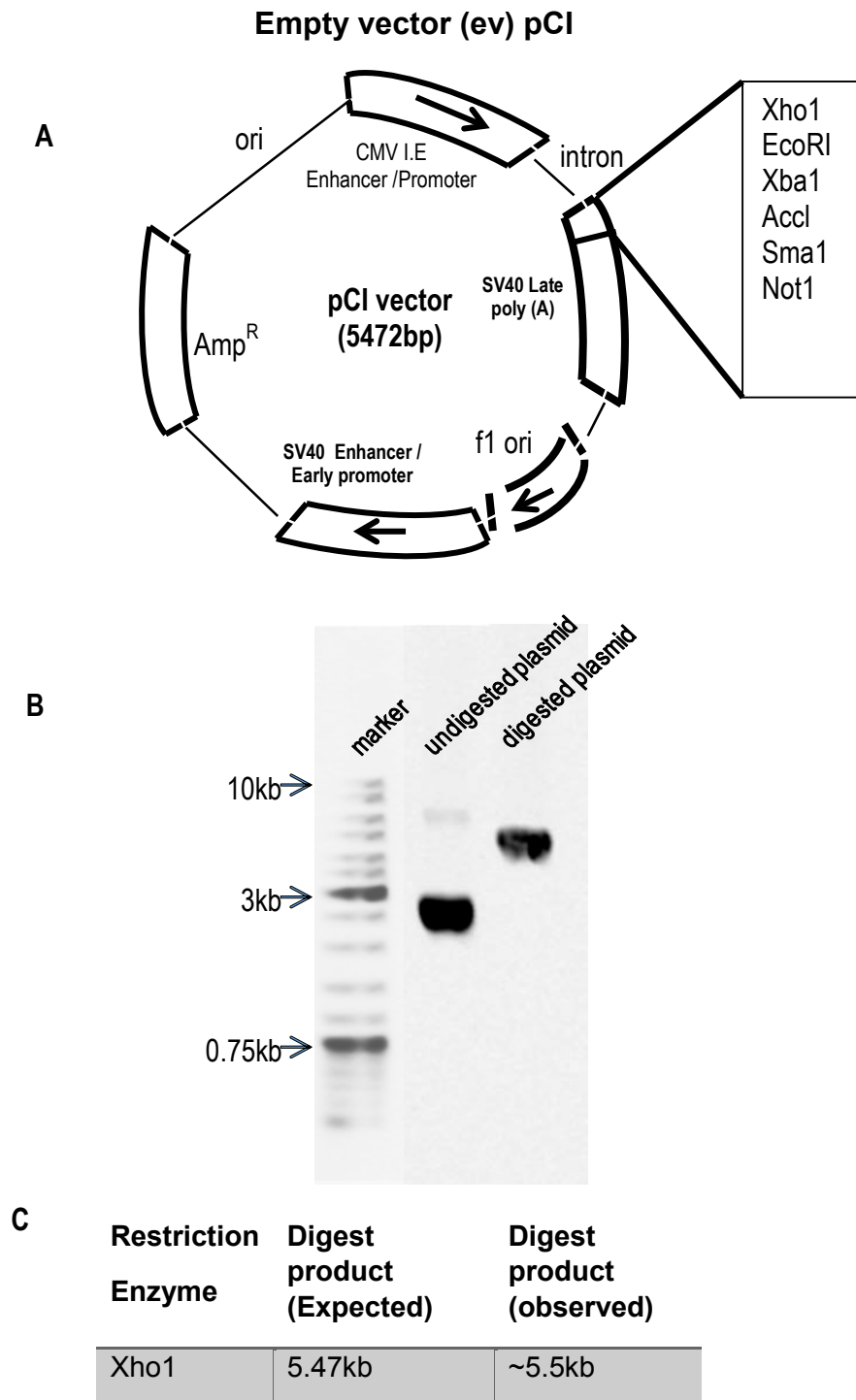
CTCF overexpression experiments were performed to assess the effect on ER expression of overexpressed CTCF. First, molecular weights (MW) of plasmid vectors used in overexpression experiments were assessed by Xho1 restriction enzyme digest to determine whether they matched the expected. Digested and undigested samples of plasmid DNA vectors were run on 1% TAE gel for 90 minutes at 120volts. The samples, impregnated with SYBR green, underwent electrophoresis and the gels viewed under ultraviolet light to detect DNA bands (see materials and methods, section 2.2.9.5). Shown in Figure 5.3A is the plasmid map for the CTCF pCI expression vector while Figure 5.3B revealed the restriction enzyme digest of the vector bearing the CTCF insert. Two bands running at >10kb and ~4kb were observed in the 'undigested plasmid lane' of the CTCF pCI gel run while the digested plasmid was observed at about 7kb in the 'digested plasmid lane'. Figure 5.3C indicates that the expected molecular weight of the digested plasmid is 6.6kb while the observed weight is ~7kb. With respect to the pCI empty vector (EV), Figure 5.4A shows the plasmid map for the pCI EV while Figure 5.4B revealed the restriction enzyme digest of the vector. There are two bands, a faint one at ~6kb and the second ~2.5kb in the ev pCI run in the 'undigested plasmid lane', while the 'digested plasmid lane' revealed a band at ~5kb in keeping with the expected weight of 5.47kb (figure 5.4C). These bands observed with the undigested plasmid are not unexpected as an uncut plasmid can assume multiple conformations ranging from the uncoiled to a supercoiled form that can migrate faster than a linear (cut) plasmid (Turner *et al.*, 2005). It will therefore reveal different molecular weights on electrophoresis. Taken together, the observed values are in line with the expected molecular weights. The ~1.5kb difference in observed molecular weight between the digested products of the CTCF pCI and EV pCI vectors reflect the additional molecular weights of CTCF and

His tag on the CTCF pCI vector. In confirming that the molecular weights of the plasmids matched the expected, results were confidently attributed to the right plasmid.



**Figure 5.3. Restriction enzyme digest of CTCF pCi expression vector.** (A) Plasmid map of CTCF pCi; (B) restriction enzyme digest gel; (C) expected and observed digest products. Molecular weights of the digested and undigested plasmid DNA are shown in (B) and the comparison of the observed and expected molecular weights of the digest product annotated in (C).



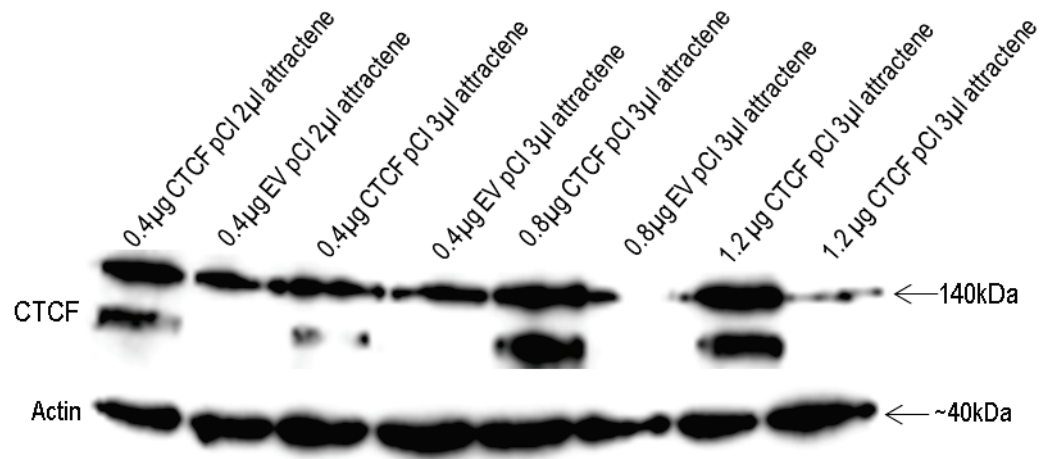


**Figure 5.4. Restriction enzyme digest of the pCI empty vector (EV).** (A) Plasmid map of pCI EV; (B) restriction enzyme digest gel; (C) Expected and observed digest products. Molecular weights of the digested and undigested plasmid DNA are shown in (B) while the comparison of the observed and expected molecular weights of the digest product is annotated (C).

### 5.6.2 Optimisation of CTCF plasmid DNA overexpression assays in MCF7 cells

To show the effect of CTCF alteration on ER $\alpha$  expression levels, CTCF plasmid DNA overexpression experiments were performed. It is essential to optimise all the steps in the transfection process as successful plasmid DNA transfection is dependent on cell type, DNA concentration, and transfection reagent type / volume and incubation periods. To determine effective plasmid DNA transfection conditions for MCF7 cells, performing multiple transfection assays using varying concentrations of plasmid DNA and transfection reagent volumes optimised CTCF overexpression assays. The incubation period post transfection remained constant at 48h according to the manufacturer's instructions.  $2.0 \times 10^4$  cells / ml of MCF7 cells were plated in each well of a 12-well plate and incubated to achieve at about 50 – 60% confluence. They were transfected with 0.4 $\mu$ g, 0.8 $\mu$ g, 1.2 $\mu$ g and 1.6 $\mu$ g of CTCF pCI and negative control pCI empty vector (EV). Varying volumes (2 $\mu$ l to 4.5 $\mu$ l) of attractene transfection reagent were used for each plasmid DNA amount and cells were incubated for 48 hours in 5% CO $_2$  at 37°C. Lysates were prepared after 48h incubation and proteins separated by SDS PAGE. Blotted membranes were probed serially with anti - CTCF (BD Biosciences, 0.5 $\mu$ g/ml), anti-ER $\alpha$  (MA 310, ThermoScientific, 5  $\mu$ g/ml), anti ER $\alpha$  (Abcam, 2647, 1 $\mu$ g/ml) and mouse monoclonal anti-actin (1:2000 dilution) antibodies. A representative panel of CTCF overexpression using 0.4 to 1.2  $\mu$ g of plasmid DNA and 2 $\mu$ l to 3 $\mu$ l of attractene is shown in figure 5.5. The figure revealed ~80% CTCF protein overexpression using 1.2  $\mu$ g of plasmid DNA and 3 $\mu$ l of attractene. Shown also are extra protein bands at ~120kDa in the lanes with overexpressed CTCF which could suggest that the extra CTCF protein produced by the cell on CTCF overexpression included isoforms with different molecular weights. There was similar overexpression level obtained with 1.6 $\mu$ g of plasmid DNA and 3 $\mu$ l of attractene (data not shown). Following CTCF overexpression, ER $\alpha$  expression was assessed using the two different anti ER $\alpha$  antibodies mentioned above and at up to three times the recommended concentrations. No ER $\alpha$  expression was

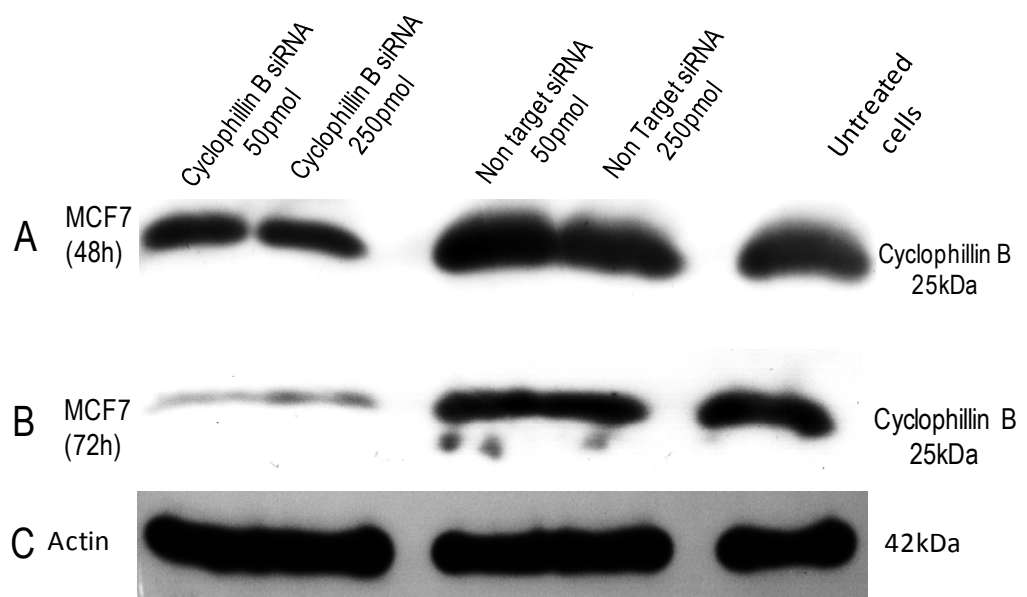
detected however at any of the CTCF transfection conditions. As protein expression was not achievable, further experiments studying gene expression were performed to determine whether CTCF overexpression had an effect on *ER* gene expression. Since higher quantities of transfected DNA can lead to unwanted activity at other gene loci (off-target effects) 1.2 µg of plasmid DNA (as opposed to 1.6 µg) transfected with 3µl of attractene and incubated for 48 hours, yielded the best transfection conditions with least potential toxicity for CTCF plasmid overexpression in MCF7 cells (figure 5.5). These conditions were used for further overexpression assays.



**Figure 5.5. Optimisation of MCF7 overexpression with a plasmid expression vector.** The effect of different amounts plasmid DNA (CTCF tagged and empty vector) and transfection reagent on CTCF protein expression is shown. Shown is about 80% CTCF protein overexpression using 1.2µg of DNA and 3µl of attractene transfection reagent compared to 1.2 µg of the empty vector (EV) pCI.

### 5.6.3 Assessment of siRNA transfection efficiency and experimental set up with the positive control cyclophilin siRNA

Assessment of the effect of CTCF on *ERα* expression involved knockdown studies with siRNA. As with overexpression studies, successful RNA interference (RNAi) experiments are dependent on cell type, siRNA concentration, and transfection reagent type / volume and incubation periods. First however, adequacy of the experimental process had to be confirmed. To test siRNA delivery and RNA knockdown efficiency in MCF7 cells, siRNA knockdown assays were performed using the positive control cyclophilin B siRNA. The *cyclophilin B* gene is a reference gene and the siRNA targeting it is validated and guaranteed to knockout the gene. The absence of knockout using this siRNA would suggest a faulty experimental set up. MCF7 cells were plated at  $1.33 \times 10^5$  cells / well in a 12 well plate and transfected with cyclophilin B siRNA at 50 pmol and 250 pmol concentrations using 2.66μl of DharmaFECT 1 as transfection reagent. Transfected cells were incubated over 48h and 72h to determine the best incubation period. As shown in Figure 5.6A, compared to the non target control, MCF7 cells revealed about 40% - 50% knockdown in cyclophilin B protein expression at 48h compared to about 80% knockdown at 72 hours noted in figure 5.6B. There is upwards of about 20-30% decrease in cyclophilin protein expression with the nontarget siRNA compared to untreated cells. Actin loading appeared uniform across the lanes. The findings confirmed that the transfection efficiency was high and that the experimental set up was optimal.



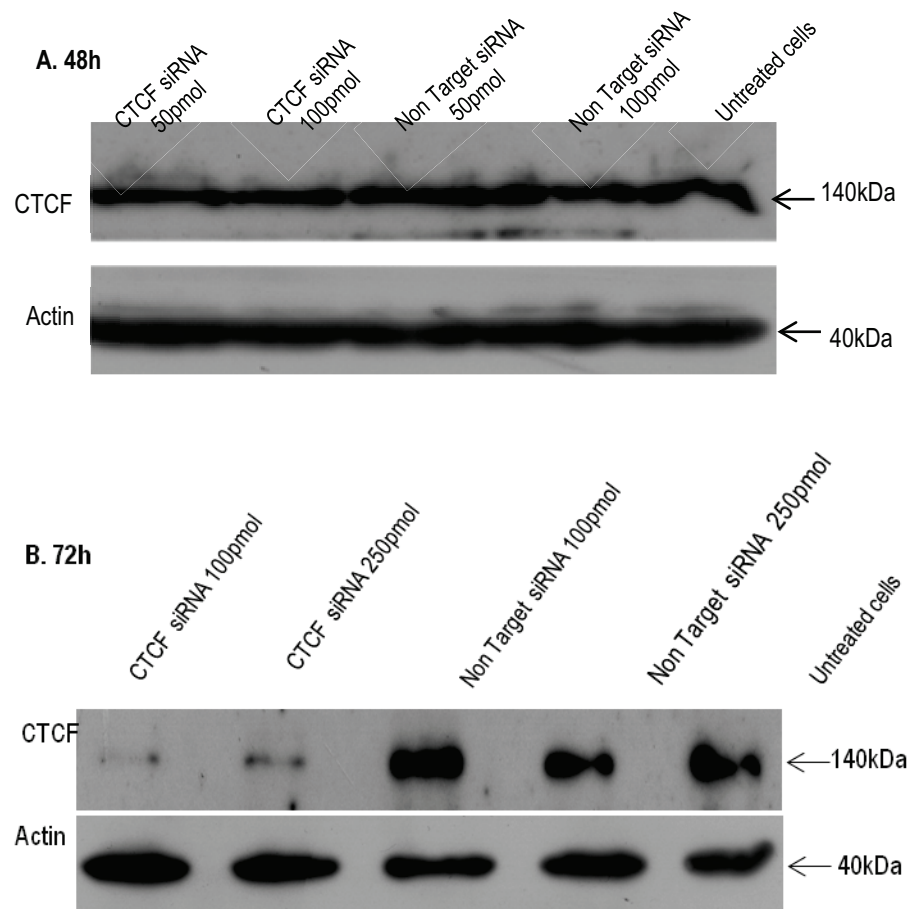
**Figure 5.6. CTCF siRNA transfection efficiency in MCF7 cells with cyclophilin B positive control.** Transfection efficiency assessed by extent of CTCF protein knockdown effect at is shown revealing about 40% - 50% CTCF protein knockdown at 48h incubation and ~80% knockdown at 72 hours incubation. Actin loading appears similar.

#### 5.6.4 Optimisation of CTCF siRNA knockdown assays in MCF7 cells

Having shown high transfection efficiency for MCF7 cells using cyclophilin siRNA and confirming the adequacy of the experimental set up (figure 5.6), siRNA optimisation experiments were performed using different CTCF siRNA concentrations and incubation periods to determine optimum siRNA transfection conditions for CTCF in MCF7 cells. Transfection reagent type and volume was limited to the manufacturer's recommendation for the MCF7 cell line. siRNA knockdown assays were performed with 50, 100 and 250 pmol respectively of CTCF siRNA and negative control non target siRNA using 2.66µl of DharmaFECT 1 transfection reagent for each well of a 12 well plate. Cells were incubated over 48h and 72h respectively. Lysates were prepared and proteins separated by SDS PAGE and blotted membranes probed serially with anti-CTCF (BD Biosciences, 0.5µg/ml dilution), anti-ER (MA 310, ThermoScientific, 5µg/ml), anti ER (Abcam 2647, 1µg/ml) and mouse monoclonal anti-actin (1:2000 dilution) antibodies. The results in figure 5.7 A, showed virtually no knockdown effect using 50pmol and 100pmol of CTCF siRNA at 48h incubation. When the experiments were repeated and incubation extended to 72h, there was about 90% knockdown of CTCF protein expression compared to the nontarget and untreated MCF7 cells (figure 5.7 B). Blotted membranes were repeatedly probed with anti-ER antibody – as also done with the overexpression studies in section 5.6.2 - to determine whether the demonstrated CTCF protein knockdown effect had any impact on ERα protein expression. Surprisingly, no bands were visualised despite using up to three times the manufacturer's recommended anti-ER antibody concentrations and with upwards of 50µg of cell lysate. This lack of visualisation was evident for both siRNA and nontarget control samples. As protein expression was not achievable due to possible technical issues with the anti ER antibodies, further experiments studying gene expression were performed to determine whether CTCF knockdown had an effect on *ER* gene expression. The combination of 250pmol of CTCF siRNA, 2.66 µl of DharmaFECT 1

transfection reagent, and  $1.33 \times 10^5$  MCF7 cell density in a single well of a 12 well plate, incubated for 72 hours, therefore served as the optimal conditions for CTCF siRNA knockdown experiments.





**Figure 5.7. MCF7 cell RNA interference (RNAi) with CTCF and non target siRNA.** (A) showing MCF7 cells transfected with 50pmol and 100pmol of CTCF and nontarget siRNA controls using 2.66 $\mu$ l of DharmaFECT1 as transfection reagent and incubated for 48h. No significant CTCF knockdown is demonstrated. (B) MCF7 cells treated with 100pmol and 250 pmol of CTCF and nontarget siRNA using 2.66 $\mu$ l of DharmaFECT1 as transfection reagent and incubated for 72h. About 90% CTCF protein knockdown with both siRNA concentrations is demonstrated.

### 5.6.5 Effect of transfection agents and reagents on MCF7 cell growth and viability

In a monolayer, progressively increasing cell confluence indicates cell growth and viability (Baydoun, 2010). Growth in turn is important in gene expression studies, as cells need to be dividing to take up nucleic material and respond appropriately to its internalisation. CTCF is essential to cell function and alterations in the expression levels could affect cell viability. To determine therefore that MCF7 cells were viable at the end of the transfection assays, cells were incubated with pCi CTCF, EV pCi, siRNA, and transfection reagents and observed over 48h to 72h. Using the transfection reagent manufacturer's recommendations, MCF7 cells were plated at  $2.0 \times 10^4$  density for overexpression studies in four wells each of a 12 well plate and incubated overnight at 37°C. On achieving 70 - 80% confluence, cells were respectively incubated with 1.2µg of CTCF pCi together with 3 µl of attractene (figure 5.8A), 1.2 µg of EV pCi with 3µl of attractene (figure 5.8B), 3 µl attractene only (figure 5.8C), and no treatment (figure 5.8D). Transfection complexes and reagents were discarded at 12 h and replaced with complete medium. They were incubated for a total of 48 hours. The results in figure 5.8 revealed considerable cell death manifested by cell rounding and loss of confluence (Baydoun, 2010) with the CTCF pCi (figure 5.8A) and somewhat less so with the EV pCi vector (figure 5.8B). There was no significant effect on cell confluence with attractene only (figure 5.8C) and about 90% cell confluence was observed with untreated MCF7 cells at 48h (figure 5.8D). It is not clear why the empty vector pCi is associated with cell loss. Ectopic expression of CTCF has previously been shown to profoundly inhibit cell growth with and without apoptosis (Rasko *et al.*, 2001; Qi *et al.*, 2003).

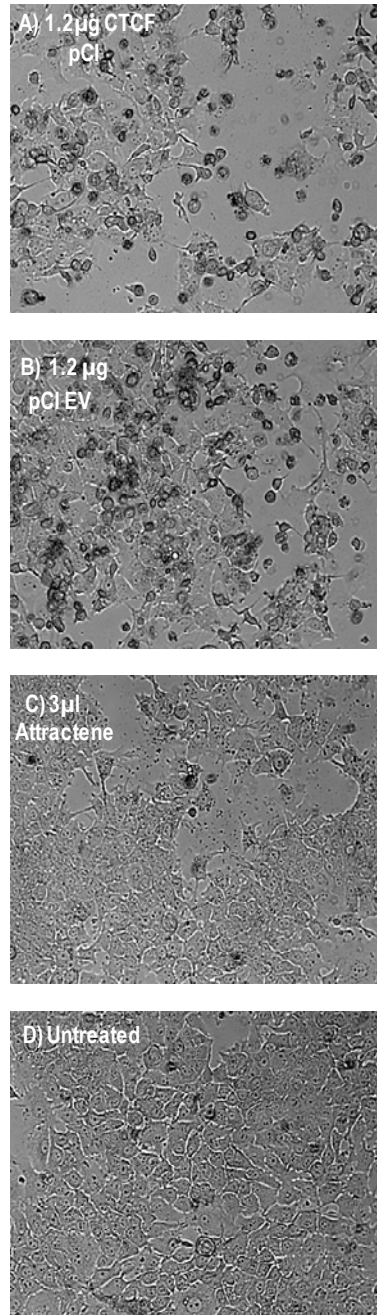
Further experiments to show the effect of CTCF siRNA and siRNA tranfection reagents on MCF7 cells were performed with cells plated at  $1.33 \times 10^5$  densities. Four wells of a 12-well plate had MCF7 cells transfected respectively with 250 pmol of CTCF siRNA

plus 2.66  $\mu$ l of DharmaECT 1 (figure 5.9A), 250 pmol of non target siRNA plus 2.66  $\mu$ l of DharmaFECT 1 (figure 5.9B), 2.66  $\mu$ l DharmaFECT1 only (figure 5.9C) and no treatment (figure 5.9D). They were incubated for 72h at 37°C and had transfection complexes and reagents replaced with antibiotic-free medium at 12h. Results shown in figure 5.9 demonstrated considerable cell death manifested by cell rounding and loss of confluence with CTCF siRNA with resultant 40% – 50% cell confluence at 72 hours (figure 5.9A). The cells incubated with DharmaFECT 1 transfection reagent and non-target siRNA only showed minimal cell rounding and about 80% confluence (figure 5.9B and C) while complete cell confluence was observed with untreated MCF7 cells at 72h (figure 5.9D).

In order to assess the viability of transfected cells still stuck to the bottom of the well, the trypan blue cell viability assay was performed. There was a 96.4% and 95.7% viability detected for MCF7 cells after 48h and 72h transfection incubation respectively. These numbers suggested that cells remained viable post transfection allowing further downstream processes.

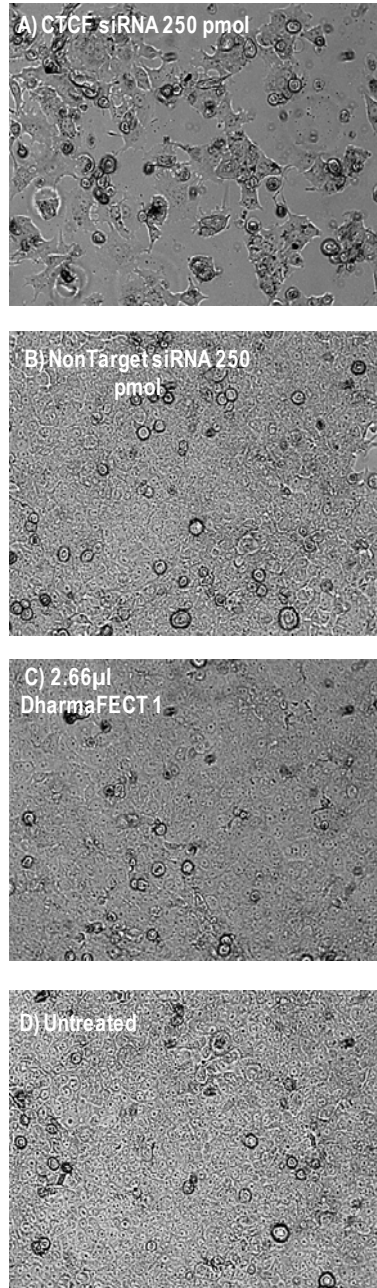
## MCF7

### Overexpression assays (48h)



**Figure 5.8. MCF7 cell response to expression vectors and attractene transfection reagent.** Shown are MCF7 cells transfected with 1.2µg of CTCF pCi plus 3 µl attractene (A), 1.2µg of EV pCi plus 3 µl attractene (B), 3 µl attractene alone (C) and no treatment (D). The panels show the extent of cell confluence and cell rounding (indicating cell death) on treatment with significant cell loss in cells transfected with plasmid expression vectors.

MCF7  
Knockdown  
assays (72h)



**Figure 5.9. MCF7 cell response to siRNA and DharmaFECT transfection reagent.** Shown are MCF7 cells transfected with 250 pmol of CTCF siRNA using 2.66 µl DharmaFECT 1 (A); 250 pmol of non target siRNA using 2.66 µl DharmaFECT 1 (B); 2.66 µl DharmaFECT 1 only (C); and no treatment (D). The panels indicate the extent of cell confluence and cell rounding (indicating cell death) on treatment.

#### **5.6.6 QPCR measurements of CTCF and ER $\alpha$ gene expression following plasmid overexpression and siRNA knockdown assays in MCF7 cells**

The previous sections of this chapter showed results obtained from CTCF overexpression and knockdown studies in MCF7 cells. The results provided optimal conditions for CTCF overexpression and knockdown in MCF7 cells. Though sufficient CTCF protein overexpression and knockdown was achieved, the effect of that change on expression levels on ER $\alpha$  protein could not be determined by western blot analysis despite the use of various concentrations of different anti-ER $\alpha$  antibodies. In order to further assess the effect of CTCF on ER  $\alpha$ , gene expression studies were performed using QPCR following transfection of MCF7 cells with plasmid expression vectors and siRNA.

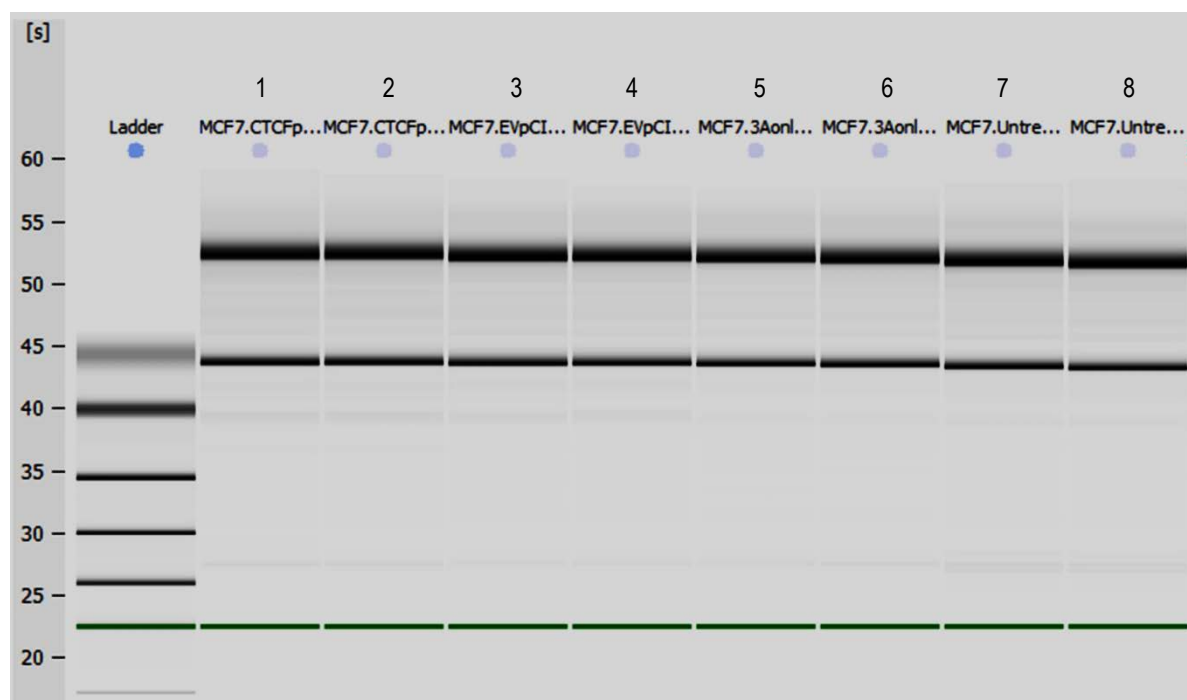
It has recently been identified that there was significant issue with the quality of some QPCR data published in the literature with respect to reproducibility and accuracy. These concerns led to the development of the Minimum Information for Publication of Quantitative Real-Time PCR experiments (MIQE) guidelines to direct and define minimum data required to publish reliable QPCR results (Bustin *et al.*, 2009). The accuracy and thoroughness of the processes upstream in the QPCR process all affect the quality and reproducibility of results. These upstream processes include RNA quantity and quality verification, primer secondary structure assessment, replicate consistency, standard curve calculation of primer efficiency, single peak melting curves of primers and linear range estimates of cDNA dilutions. The sections below described results of these experiments starting from the quality of RNA extracted from CTCF overexpressed and knockdown MCF7 cells through assessment of the adequacy of primer structure, melting points, standard curves to the actual QPCR experiments.

#### 5.6.6.1 Bioanalyser and Nanodrop estimation of RNA quality and concentrations

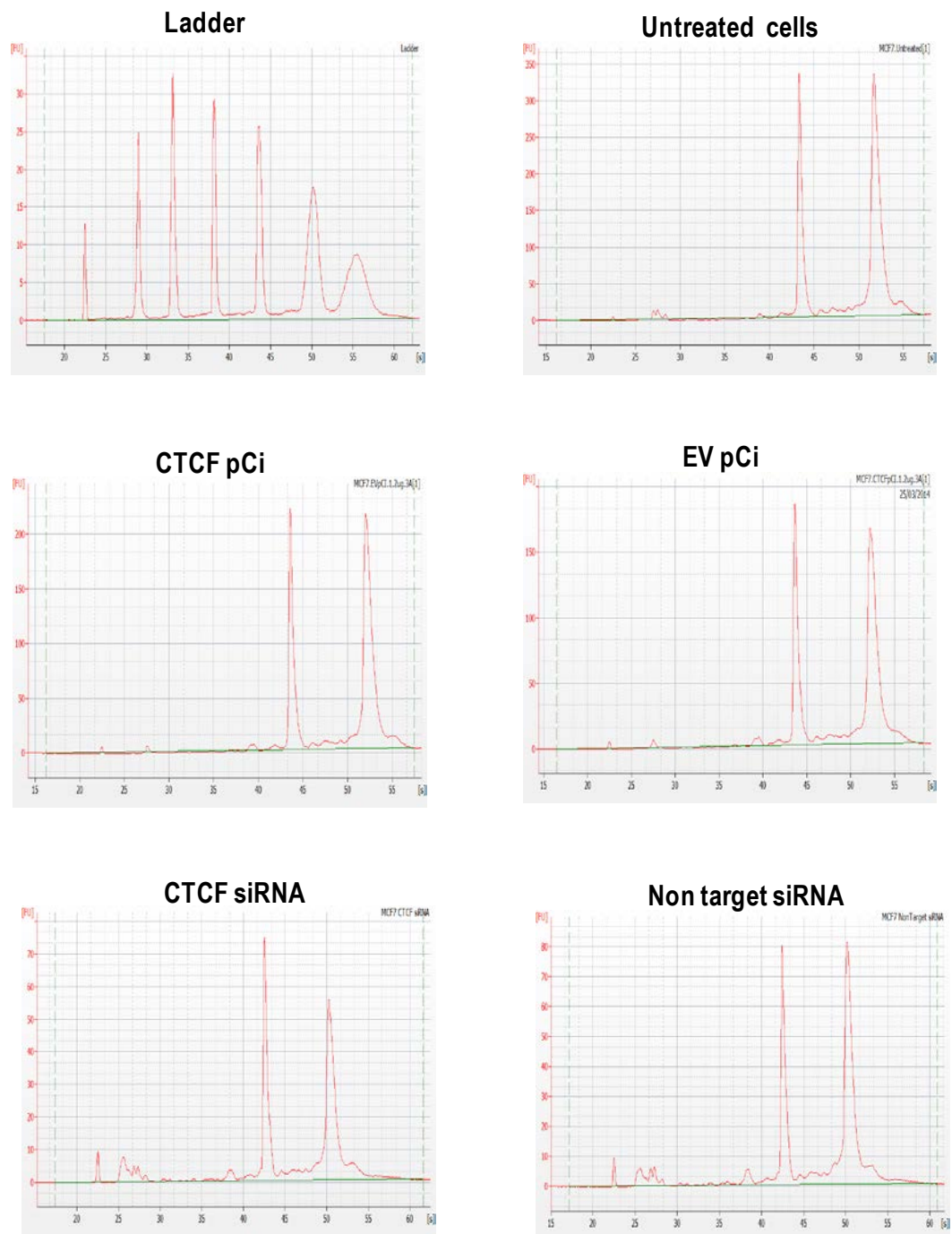
RNA quality is one of the many determinants of reproducible and dependable QPCR assays. RNA is extremely fragile, degrades easily and is also readily contaminated (Fleige and Pfaffl, 2006; Bustin *et al.*, 2009). Using optimised conditions for CTCF overexpression and siRNA transfection summarised in sections 5.6.2 and 5.6.4, MCF7 cells were transfected. In order to determine CTCF and ER $\alpha$  mRNA expression following plasmid overexpression and siRNA knockdown by QPCR, transfected cells first underwent RNA extraction, were treated with DNase to remove possible contaminating genomic DNA and analysed for RNA concentration and quality. Aliquots of 1.5  $\mu$ l and 1  $\mu$ l of extracted RNA were used to determine concentration and quality according to the manufacturer's instructions using the Nanodrop spectrometric machine and Agilent Bioanalyser hardware respectively. As shown in figure 5.10, using the Agilent bioanalyser two clear bands that represent the 18S and 28S ribosomal proteins were apparent. There were no extra bands suggesting that there were no contaminating nucleic acids in the samples. Figure 5.11 showed the electropherograms from the bioanalyser and confirmed the 18S and 28S RNA by having only two peaks in all the samples. The electropherograms are generated by the bioanalyser by plotting the intensity of fluorescence with size / migration time of the RNA in each sample. The bioanalyser was also used to estimate the RNA integrity number (RIN) of the samples. RNA integrity number (RIN) of 8 and over is taken as good quality with a value of 10 the highest possible quality (Schroeder *et al.*, 2006). Table 5.1 showed that the lowest RIN in the extracted RNA samples was 9.4 in the knockdown sample with CTCF siRNA. Indeed the untreated MCF7 cell sample achieved the maximum of 10. This result suggested that RNA used for the QPCR experiments were of pure quality with no significant contamination or degradation. This conclusion was further confirmed by the 260 / 280 absorbance ratio determined by the Nanodrop spectrometer (table 5.1) where the lowest absorbance ratio for this study was 1.93 in the sample with EV pCi. A reading

of between 1.8 and 2.0 is considered a pure nucleic acid material (Fleige and Pfaffl, 2006). Finally, RNA concentrations were estimated using the Nanodrop spectrometer and are also listed in table 5.1.





**Figure 5.10. Densitometry plot of RNA quality assessment using the Agilent Bioanalyser.** Representative Figure showing RNA quality assessment performed with the Agilent Bioanalyser. CTCF pCI (lanes 1 and 2), EV pCI (3 and 4), attractene only (lanes 5 and 6) and untreated cells (lanes 7 and 8) were assessed for RNA quality. 1µl of extracted RNA for each treatment was subjected to analysis using the Agilent Bioanalyser and samples were loaded in duplicate (see materials and methods, section 2.2.8.2). Sample lanes 1 to 8 show only two clear bands that represent the 18S and 28S ribosomal protein fractions. A lane with only two bands indicate pure uncontaminated and un-degraded RNA sample. The green band at the bottom of all the lanes represents an internal standard used to align data from the ladder and the sample lanes. The alignment compensates for drift effects associated with a chip run.



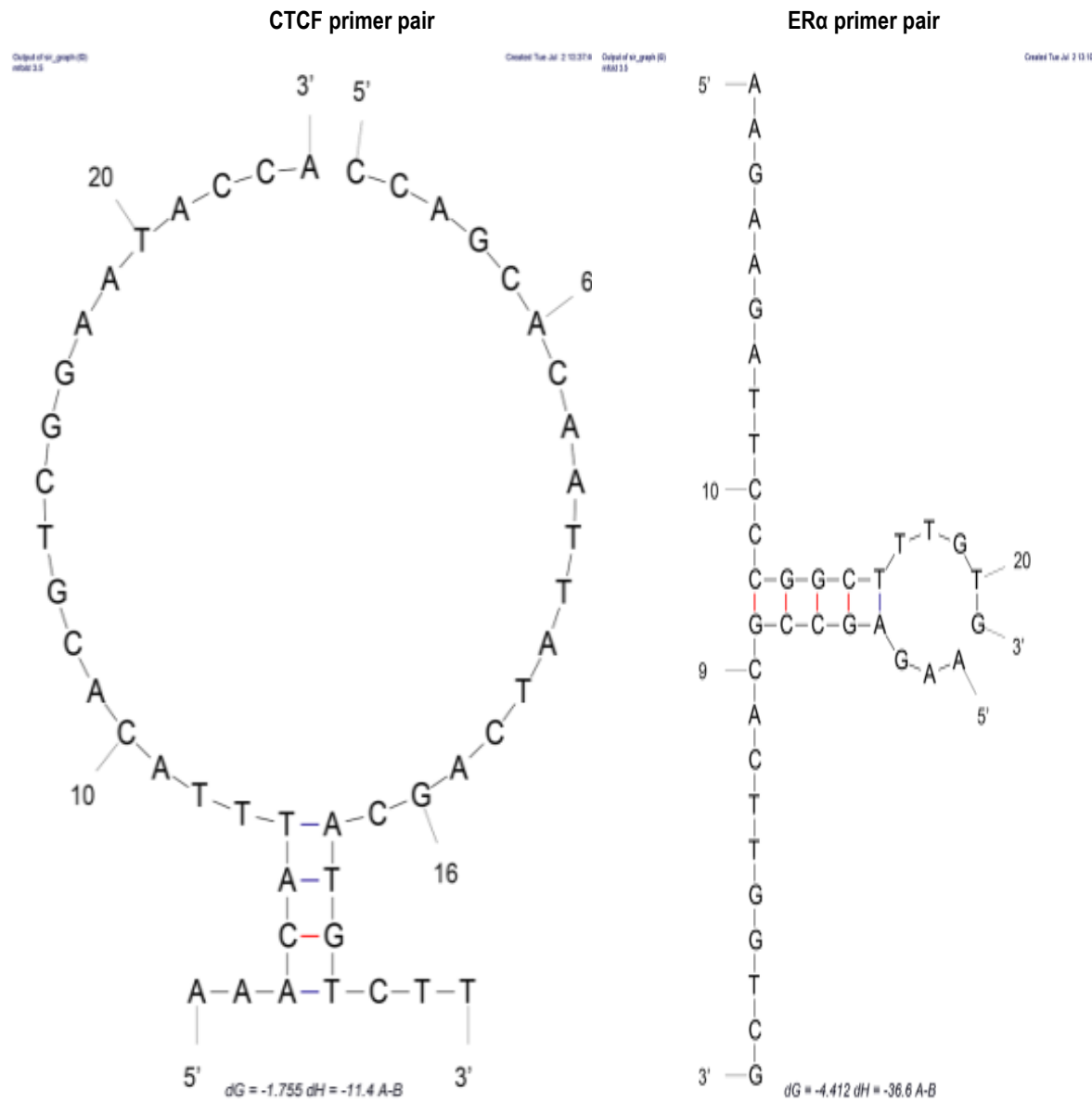
**Figure 5.11. Spherograms of RNA quality assessment using the Agilent Bioanalyser.** Spherograms of RNA quality assessment of treated (CTCF pCi, EV pCi, CTCF siRNA and non target siRNA) and untreated MCF7 cells shown in Figure 5.10 are represented here. The RNA fragment peaks (first peak from the left is 18S and second peak 28S) in the graphs represented the ribosomal bands noted in the densitometry plot of Figure 5.10.

Sample identity	RNA concentration	260 / 280 ratio	RIN
	ng / $\mu$ l	Nanodrop	Bioanalyser
	Nanodrop		
CTCF pCi	840	1.94	9.90
EV pCi	881	1.93	9.80
CTCF siRNA	355.4	1.97	9.40
Non target siRNA	360.6	1.98	9.90
Untreated MCF7 cells	880.9	1.99	10.0

**Table 5.1. Bioanalyser and Nanodrop spectrometric estimation of RNA concentrations and purity.** Bioanalyser estimate of RNA quality is shown by the RNA integrity number (RIN) for each sample. Also shown is the Nanodrop estimation of RNA purity and RNA concentration. A pure RNA sample is acceptable as having a 260 / 280 OD ratio of 1.7 to 2 and a RIN of at least 8 out of 10.

#### **5.6.6.2 DINAmelt<sup>R</sup> prediction of secondary structures of primer pairs for QPCR**

Correct primer annealing is essential for QPCR processes and efficient primer amplification can be prevented by extended nucleic acid secondary structures (Shiple, 2013). In order to determine the extent of secondary structures in the primers (kindly generated by F Docquier, Essex University) used for QPCR, two state melting (hybridization) prediction was performed using the DINAmelt software. The software assesses nucleic acid interactions and predicts melting, folding and hybridisation. Forward and reverse sequences of all primers used for QPCR were entered into the Two-state melting (folding) section of the DINAmelt software and assessed at temperature = 55C, magnesium = 5nH, sodium (Na) = 50nH for RNA folding pattern (Markham and Zuker, 2005). As shown in figures 5.12 and 5.13, none of the primers possessed an exaggerated secondary structure. This result predicted that there would be minimal primer interference in qPCR assays.



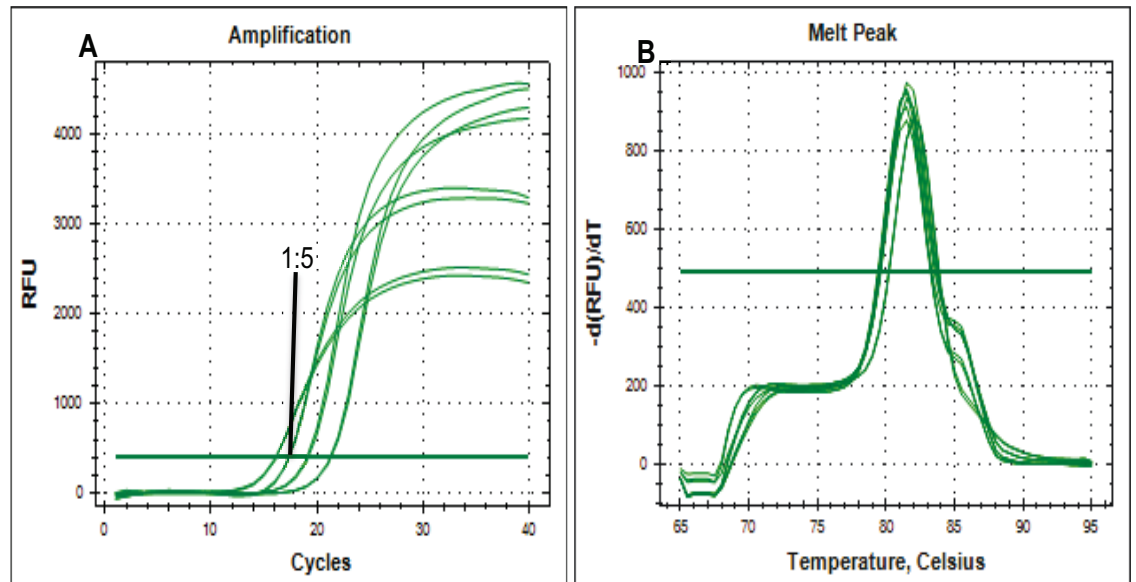
**Figure 5.12. Nucleic acid primer pair fold pattern for CTCF and ER $\alpha$  primer pairs.**

Forward and reverse primers for CTCF and estrogen receptor (ER)  $\alpha$  primers were assessed for fold pattern using the DINAmelt<sup>R</sup> software. The simulated fold pattern was assessed at a temperature of 55C with 5nM magnesium level and sodium at 50nM. The predicted secondary protein structures are shown and indicate that there is minimal secondary structure formation between the primer pairs.



### 5.6.6.3 QPCR assay optimisation: linear range and dilution factor for cDNA

The level of cDNA dilution for QPCR amplification is important as it determines the shape of the amplification curve and has an impact on Cq values. The Cq level refers to the cycle number at which the doubling of cDNA template goes past a threshold and enters the logarithm phase. If a cDNA sample is too dilute, it could give spuriously high Cq values as the cDNA template concentration would be iatrogenically low and would need more doubling cycles to achieve a level that would go past the threshold and enter the logarithm phase. To achieve uniformity, the same dilution factor was applied across all cDNA samples. To determine the best cDNA dilution for qPCR assays, the linear range (cDNA dilution factor) for MCF7 cells was therefore assessed. Untreated MCF7 cDNA reverse transcribed from 1µg of RNA was serially diluted to achieve 1:1, 1:5, 1:25 and 1:125 dilutions. A GAPDH primer pair was used for amplification. In the amplification plot shown in figure 5.14, there was a sequential increase in Cq value indicating a progressively diminishing amount of starting material as expected with increasing sample dilution. The shape of the amplification curve determines the best cDNA dilution factor and as seen in figure 5.14, the 1:5 dilution represented the best fit as samples which are too concentrated have S shapes while the overdilute samples may have no shape at all. A dilution factor of 1:5 for all cDNA samples was therefore used for the QPCR experiments. The melt peak curves in figure 5.14 revealed single peaks indicating that the QPCR assay demonstrated specificity. Table 5.2 showed that the change in Cq ( $\Delta Cq$ ) value for respective fold dilution was 1.21, 1.78 and 2.16. A perfect biological system with 5-fold dilutions would show a consistent  $\Delta Cq$  of 2.23 ( $2^n=5$ ).



**Figure 5.14. Testing primers: amplification plots for linear range.** MCF7 cDNA (reverse transcribed from 1 $\mu$ g of RNA) was amplified by QPCR using a GAPDH primer pair. Samples were serially diluted from 1:1 to 1:5, 1:25, and 1:125. QPCR was performed using kapa mastermix and each well in the 96-well plate was loaded with 5  $\mu$ l samples. Experiments were performed in duplicate and on three different occasions. (A) The change in Cq values (cycles on the x-axis in the amplification plot) with increasing dilution is shown in the amplification plot. The amplification plot representing 1:5 dilution is identified with a black line and the Cq is the cycle number at which it intersects the horizontal threshold line and is 17.31. (B) The melting plot shows single peaks confirming the absence of contaminating nucleic material.



Cell line	Primer	Dilution factor	Cq values (duplicate)	Cq mean $\pm$ SD	Melting temp (C)
<b>MCF7</b>	GAPDH	1:1	16.05	16.10 $\pm$ 0.07	82.00
<b>MCF7</b>	GAPDH	1:1	16.15		
<b>MCF7</b>	GAPDH	1:5	17.30	17.315 $\pm$ 0.02	81.50
<b>MCF7</b>	GAPDH	1:5	17.33		
<b>MCF7</b>	GAPDH	1:25	19.03	19.095 $\pm$ 0.092	81.50
<b>MCF7</b>	GAPDH	1:25	19.16		
<b>MCF7</b>	GAPDH	1:125	21.21	21.255 $\pm$ 0.063	81.50
<b>MCF7</b>	GAPDH	1:125	21.30		

**Table 5.2. Tabular annotation of Cq values of amplification plots for linear range.**

The Table shows Cq values expressed as mean  $\pm$  SD together with the difference in Cq values between duplicates as shown in figure 5.14. The narrow differences in duplicates suggested accurate pipetting. Also shown is the single DNA melting DNA temperature at different cDNA dilutions indicating that the nucleic acid material was not contaminated.

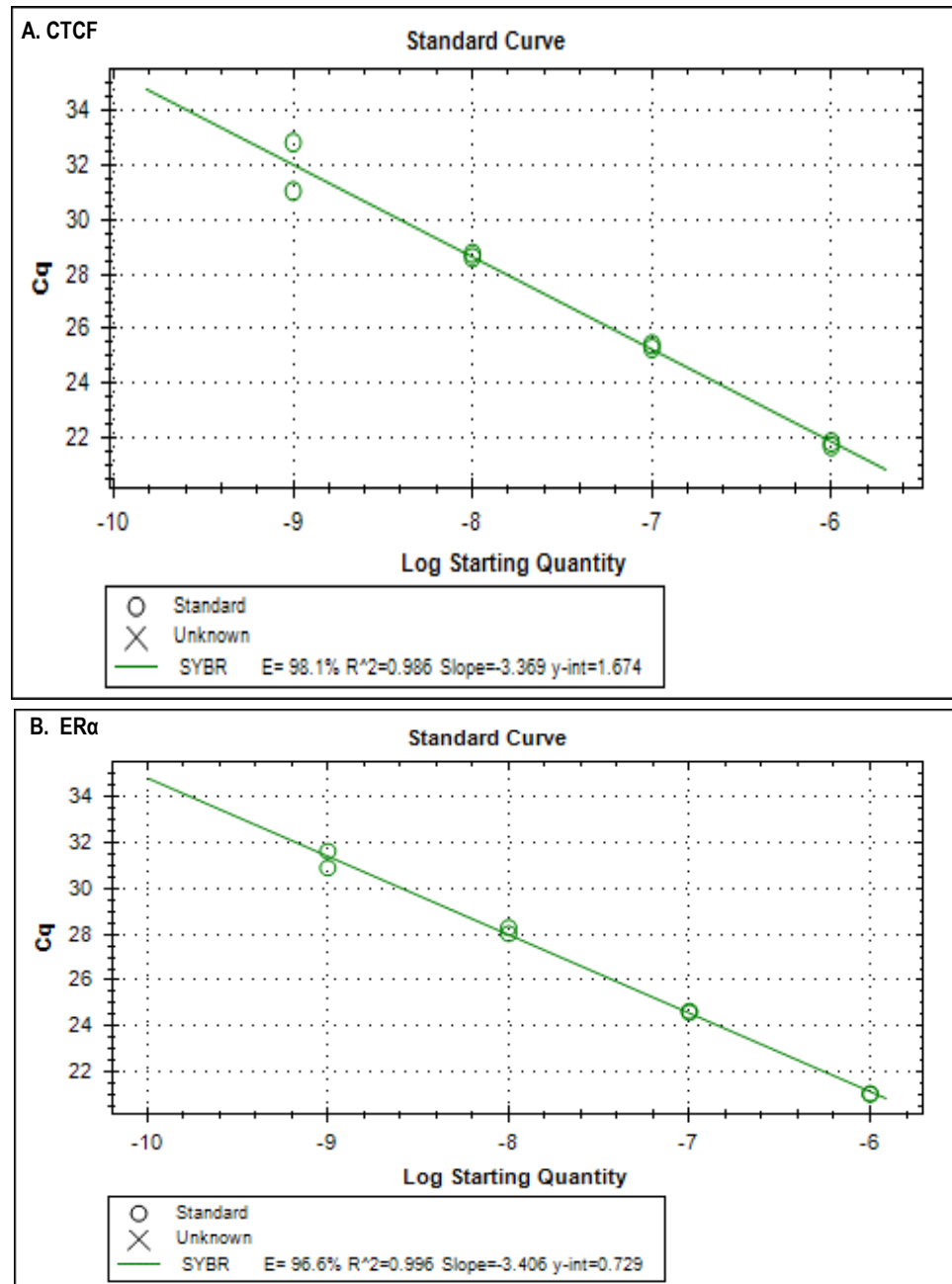
#### 5.6.6.4 QPCR optimisation: standard curves and melting curves.

An accurately validated QPCR assay is determined by the combination of a linear standard curve with an amplification efficiency of 90 – 105%, slope of -3.9 to -3.0 and  $R^2$  value  $>0.98$  (all influenced by replicate consistency) and single peak melting curves (Bustin *et al.*, 2009). The standard curve is a tool used to assess efficiency and sensitivity of QPCR assays. In effect, it tests the ability of the QPCR set up to consistently double nucleic material with each temperature cycle. Efficiency is mathematically calculated from the slope of the curve. The  $R^2$  value assessed the fit between experimental data and the regression line of the standard curve and shed light on the variability between sample replicates. It also gave an indication as to whether different starting template numbers possessed similar amplification efficiency.

The standard curve for QPCR experiments performed in this thesis was determined using serial log dilutions (1:1, 1:10, 1:100, and 1:1000) of untreated MCF7 cDNA. Using CTCF, ER $\alpha$ , GAPDH and TBP primer pairs the associated cDNA was amplified in untreated MCF7 cells. Figure 5.15 showed that the standard curve for the CTCF primer pair (A) was linear and demonstrated a slope of -3.369,  $R^2$  of 0.986 and efficiency of 98.1%. The ER $\alpha$  primer pair demonstrated a slope of -3.406,  $R^2$  of 0.996 and efficiency of 96.6%; GAPDH primer pair had a slope of -3.343,  $R^2$  of 0.996 and an efficiency of 95.5%; and TBP primer pair showed a slope of -3.367,  $R^2$  of 0.987 and efficiency of 98.2%. These values were within expected limits of an accurately optimised QPCR assay. Further inspection of  $C_q$  values for each log dilution of the GAPDH primer pair, revealed absolute  $C_q$  values of 18.5, 22.25, 25.89 and 28.73 respectively for the neat, 1:10, 1:100 and 1:1000 dilutions. These values represented  $C_q$  difference of 3.75, 3.64 and 2.82 between the log dilutions. Perfect biological systems would generate a constant  $C_q$  difference of 3.2 ( $2^n=10$ ) between respective log dilutions (10 fold). As previously observed in section 5.6.6.3 using 5-fold dilutions, the results shown here

however confirmed that the fidelity for each replication is not consistently 100% perfect for each cycle emphasising that these systems are not perfect. Values obtained with the slope, efficiency and  $R^2$  in the standard curve which all fell within normal limits in this thesis represented a rigorously and accurately optimised QPCR assay. Standard curves were subsequently performed for each experimental sample that underwent QPCR.

Further validation of the reliability of the QPCR assays was provided by the peak melting point. This factor described the temperature at which double stranded DNA (dsDNA) dissociates into single stranded DNA (ssDNA) and is associated with a fall in fluorescence. Nucleic acids of different lengths and sequences produce different peak melt points. Therefore a single melt peak indicates the presence of a specific nucleic acid product from a specific primer pair and therefore tests the specificity of the QPCR reaction. The results in the melt peak plots of figure 5.16 confirmed that only the specific CTCF, ER $\alpha$ , TBP and GAPDH cDNA were amplified by their respective primer pairs and that there was no contamination with other nucleic acid material. Taken together, the results shown in Figures 5.15 and 5.16 indicated that the primers, QPCR hardware, cDNA dilution factor and pipetting accuracy for the set of QPCR experiments were of optimum standard and met MIQE guidelines (Bustin *et al.*, 2009).



**Figure 5.15. Standard curves for QPCR efficiency.** MCF7 cDNA was reverse transcribed from 1µg of RNA and serially diluted 10 fold from neat to 1:10, 1:100 and 1:1000. Samples were amplified by QPCR with (A) CTCF, (B) ERα and (C) GAPDH and (D) TBP primer pairs using Kapa mastermix. The reaction efficiency, slope and R<sup>2</sup> values for each primer is indicated in the box for each graph. The values were within expected ranges for an accurately optimised QPCR assay.

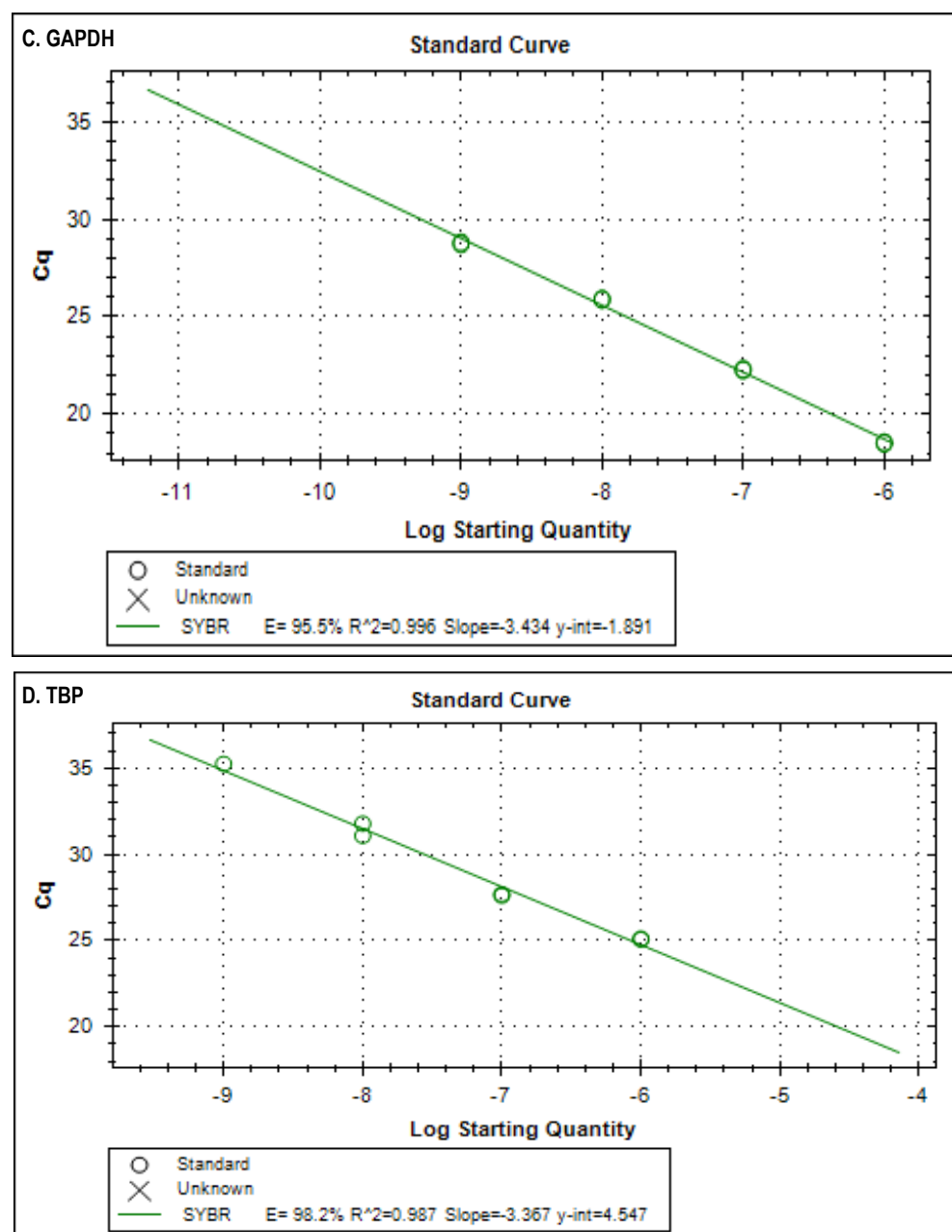
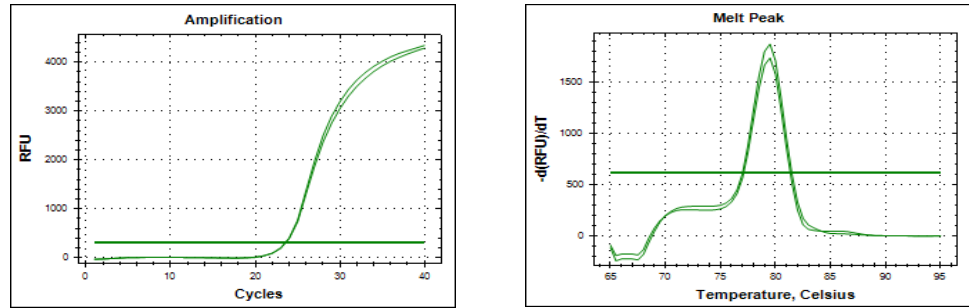
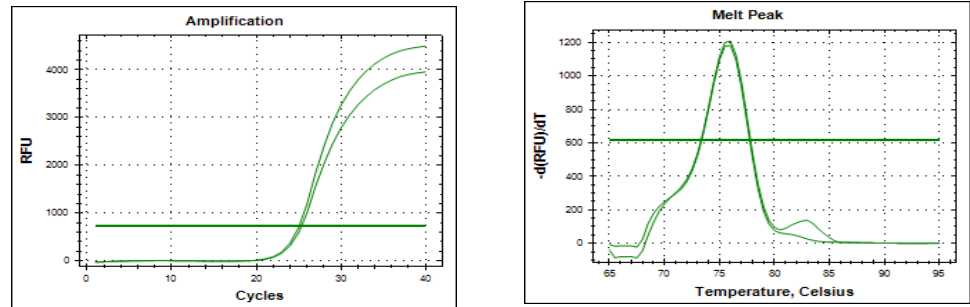


Figure 5.15 continued

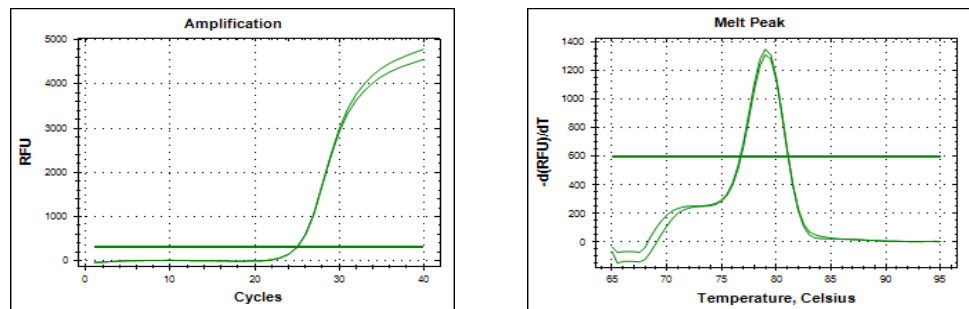
## CTCF



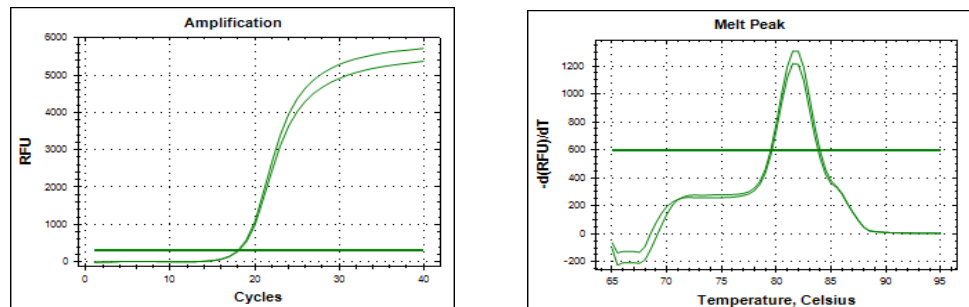
## Estrogen Receptor $\alpha$ (ER $\alpha$ )



## TBP



## GAPDH



**Figure 5.16. Assessment of QPCR replicate consistency and melting curves.** MCF7 cDNA (reverse transcribed from 1 $\mu$ g of RNA) at 1:5 dilution was amplified in duplicates by QPCR using CTCF, estrogen receptor (ER)  $\alpha$ , TATA box protein (TBP) and GAPDH primer pairs respectively. The amplification curve indicating replicate consistency for each primer pair is shown. Also shown is the single peak melting point for each primer pair indicating the absence of nucleic acid contamination.

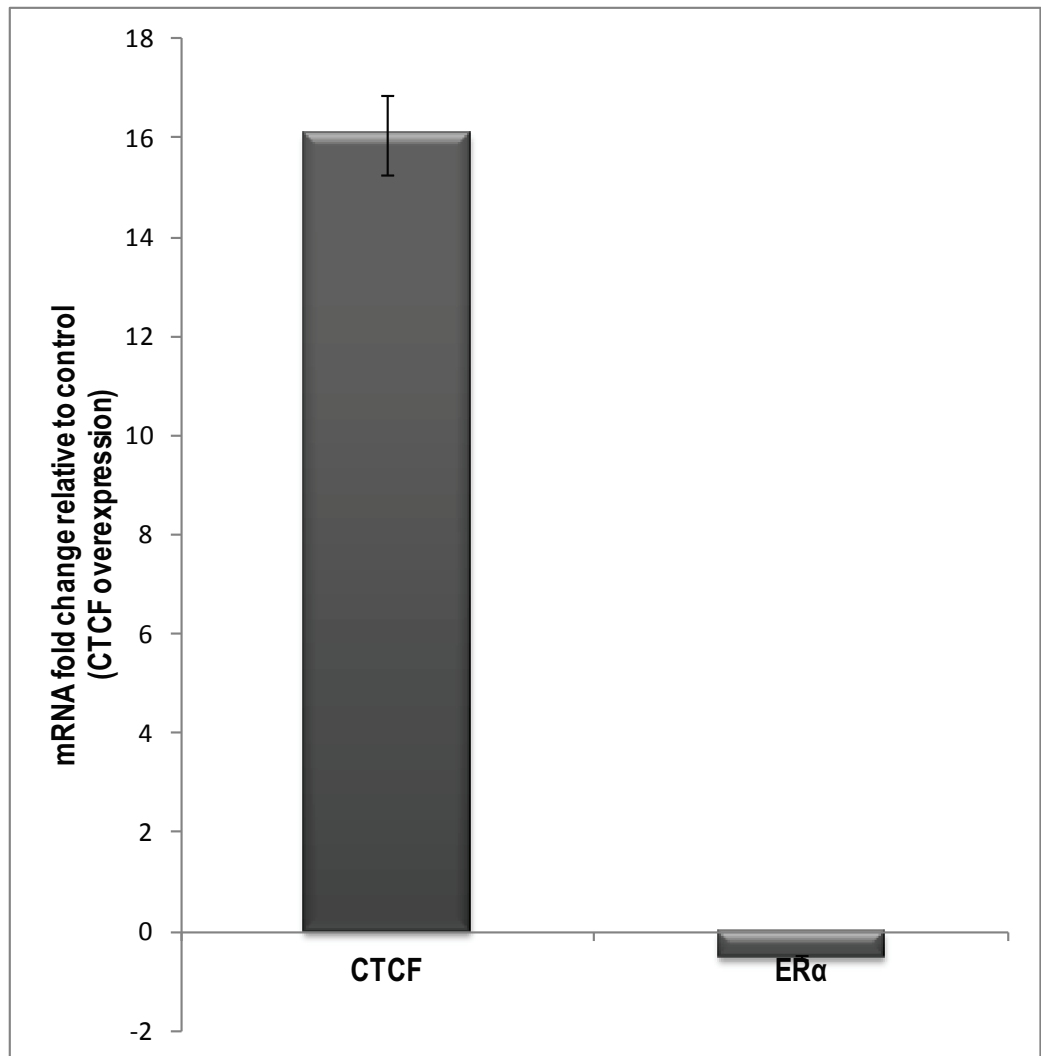
#### **5.6.6.5 Variation in CTCF mRNA expression and ER $\alpha$ expression response in MCF7 breast cancer cells**

As mentioned by Ross-Innes *et al.* (2011) and suggested by the mass spectrometry results in chapter 3 of this thesis, CTCF appeared to possess greater interaction in ER positive breast cancer cell lines compared to ER negative breast cells. While higher levels of exogenous estrogen downregulated CTCF expression (Del Campo *et al.*, 2014), it is not clear what direct effect changes in *CTCF* gene expression levels may have on *ER $\alpha$*  gene expression in ER positive breast cancer cell lines. In order to determine the impact of *CTCF* gene expression level on *ER $\alpha$*  expression, MCF7 cells (ER / PR positive) were subjected to CTCF plasmid overexpression and siRNA knockdown. For overexpression assays, MCF7 cells were transfected with the expression vector CTCF pCI and an empty vector negative control (ev pCI). cDNA was transcribed from 1ug of RNA extracted from overexpressed MCF7 cells and subjected to QPCR. The results in table 5.3 and figure 5.17 showed that CTCF overexpression fold change (that is, the difference between CTCF pCI and EV pCi) of 16.08 was associated with a fold change of -0.52 in *ER $\alpha$*  gene expression. Further experiments were performed with CTCF siRNA and nontarget siRNA controls on MCF7 cells, which were incubated for 72 hours. The results in tables 5.4 and figures 5.18 showed that a CTCF knockdown fold change of 4.54 was associated with just -0.029 fold change in *ER $\alpha$*  mRNA expression level. The expression of reference genes GAPDH and TBP remained constant for both test material and controls in each set of experiments which indicated that the change(s) identified in *CTCF* and *ER $\alpha$*  expression were not due to variations in the general expression of RNA in the cells.

	<b>Cq <math>\pm</math>SD CTCF pCI</b>	<b>Cq <math>\pm</math> SD EV pCI</b>	<b>Cq change</b>	<b>Fold change</b>
<b>CTCF</b>	<b>17.005 <math>\pm</math> 0.0495</b>	<b>21.015 <math>\pm</math> 0.0353</b>	<b>4.01</b>	<b>16.08</b>
<b>ER<math>\alpha</math></b>	<b>21.5 <math>\pm</math> 0.071</b>	<b>20.78 <math>\pm</math> 0.0283</b>	<b>-0.72</b>	<b>-0.52</b>
<b>GAPDH</b>	<b>17.425 <math>\pm</math> 0.106</b>	<b>17.175 <math>\pm</math> 0.064</b>	<b>0.25</b>	
<b>TBP</b>	<b>23.34 <math>\pm</math> 0.14</b>	<b>22.745 <math>\pm</math> 0.191</b>	<b>0.59</b>	

**Table 5.3: QPCR of plasmid vector overexpression assays, Cqs and fold change.** Tabular annotation of Cq levels for mRNA expression and fold changes for CTCF and estrogen receptor (ER $\alpha$ ) genes on MCF7 cells treated with CTCF expression vectors (CTCF pCI and empty vector pCI). The Cqs for reference genes GAPDH and TBP are also noted and the Cq change indicated whether the Cq difference between test sample (CTCF pCI) and control (EV pCI) was due to possible changes in cellular protein levels. A Cq change of less than 0.8 in reference genes is acceptable as no difference in gene expression levels.



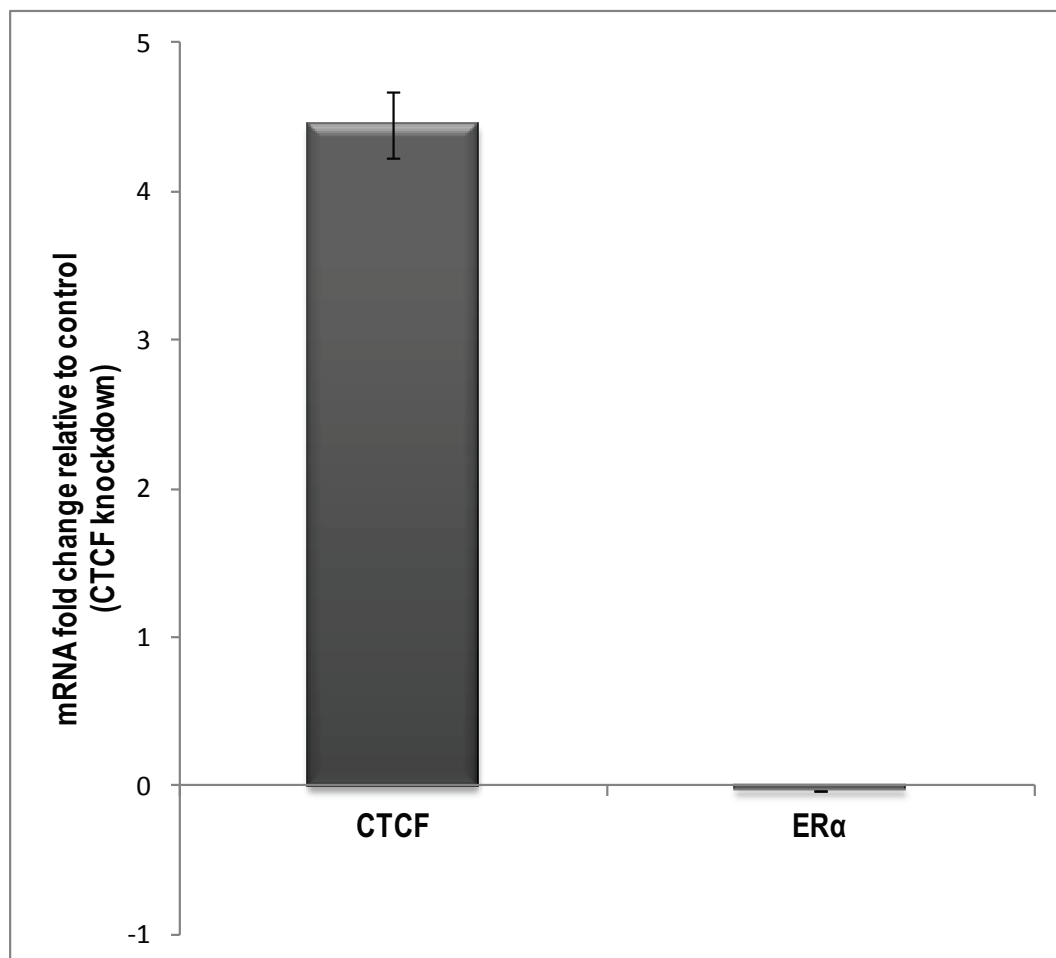


**Figure 5.17. CTCF and ERα mRNA fold change on CTCF overexpression.**

MCF7 cells were overexpressed with a CTCF expression vector (CTCF pCI) and control empty vector (EV) pCI. cDNA was reverse transcribed from 1µg of extracted RNA and subjected to QPCR. Values are the average of two experiments, performed in duplicate. Bar diagram shows *CTCF* fold change compared to (*ER*)  $\alpha$  indicating no significant change in *ER* $\alpha$  expression.

	<b>Cq ± SD</b> <b>CTCF siRNA</b>	<b>Cq ± SD</b> <b>Non target</b> <b>siRNA</b>	<b>Ct change</b>	<b>Fold change</b>
<b>CTCF</b>	24.06 ± 0.014	21.945 ± 0.078	2.11	4.45
<b>ERα</b>	20.835 ± 0.219	21.005 ± 0.007	-0.17	-0.029
<b>GAPDH</b>	18.45 ± 0.1697	18.52 ± 0.0567	0.07	
<b>TBP</b>	23.68 ± 0.106	23.455 ± 0.0778	0.22	

**Table 5.4: Cq values and fold change for CTCF siRNA knockdown assays.** Tabular annotation of Cq levels for mRNA expression and fold changes for CTCF and estrogen receptor (ER) on MCF7 cells treated with CTCF siRNA and Non target siRNA. The Cqs for reference genes GAPDH and TBP indicated whether the Cq difference between test sample (CTCF pCI) and control (EV pCI) was due to possible changes in cellular protein levels. A Cq change of less than 0.8 in reference genes is acceptable as no difference in gene expression levels.



**Figure 5.18. CTCF and ER mRNA fold change on siRNA knockdown.** MCF7 cells were transfected with CTCF siRNA and non target siRNA control for siRNA-mediated knockdown. cDNA was reverse transcribed from 1µg of extracted RNA and subjected to QPCR. Values are the average of two experiments, performed in duplicate. Bar diagram shows *CTCF* fold change compared to *estrogen receptor (ER) α*.

## CHAPTER 6

---

### Discussion – Regulatory effect of CTCF on ER $\alpha$ expression

There is a growing body of evidence suggesting a connection between CTCF and ER $\alpha$  with some publications reporting that CTCF and ER colocalised at estrogen response elements and that they both recruited similar corepressors such as HDAC to suppress gene transcription (Lutz *et al.*, 2000; Kawai *et al.*, 2003; Ross-Innes *et al.*, 2011). There also appeared to be a certain bias of CTCF interactions for ER positive breast cancer cells as opposed to the ER negative ones (Ross-Innes *et al.*, 2011). This latter finding was supported by results described in chapter 3 of this thesis. This chapter therefore hypothesized that CTCF may directly regulate the function of ER $\alpha$  in ER-positive breast cancer cells. CTCF protein and gene expression levels were assessed following transfection with an overexpression vector carrying CTCF or by silencing CTCF gene expression using siRNA. Western blot analysis and QPCR were employed to monitor the effect of CTCF overexpression or knockdown on ER $\alpha$  expression at the mRNA and protein levels. If changes were observed, chromatin immunoprecipitation (ChIP) assays using anti-CTCF antibodies were to be performed and amplification of *ER* and *PR* genes would then be tested to check whether CTCF proteins bound directly to *ER* and *PR* promoters *in vivo*. If demonstrated, identification of the precise CTCF binding sites on *ER* and *PR* gene promoters was to be investigated by electrophoretic mobility shift assay (EMSA).

The aim of this chapter was to determine protein and gene expression to establish whether CTCF regulated ER $\alpha$  expression. CTCF protein expression was readily demonstrated in all batches of untreated and transfected whole – cell MCF7 lysates by western blot. It was therefore somewhat surprising that while protein expression levels for ER $\alpha$  were determined for whole cell lysates as shown in figure 3.1, the same was not reproducible when cells underwent transfections. Anti ER antibodies obtained from

Abcam and ThermoScientific were used in this series of experiments. The monoclonal antibody to ER $\alpha$  obtained from Abcam was raised against a synthetic peptide corresponding to estrogen receptor alpha, amino acid 247-261, internal sequence EVGMMKGGIRKDRRG while the anti ER $\alpha$  antibody from ThermoScientific was also raised against a synthetic peptide corresponding to the residues E (247) V G M M K G G I R K D R R G (261) of the ER $\alpha$  DNA binding domain. The manufacturers confirmed that the antibodies were able to detect the corresponding antigen in MCF7 whole lysates a claim that was confirmed in figure 3.1. Though the same quantity of cell lysate (25 $\mu$ g - 50 $\mu$ g) recommended by the manufacturers was tested, the anti-ER $\alpha$  antibodies were not able to detect ER $\alpha$  protein on western blots after cells were transfected. The reasons behind this finding are not clear but could relate to poor antibody quality. Antibody quality however, would not fully explain the disparity in ER $\alpha$  detection between untreated and transfected whole cell MCF7 lysates as the antibodies detected ER $\alpha$  in untreated whole MCF7 lysates. It is interesting that transfection can result in unexpected morphologies and abnormalities in target cells and a possible result of those changes could be epitope masking. Masking of an epitope occurs due to a change in the conformation or electrostatic charge of an antigen, a situation that could supervene when cells are transfected. In support of epitope masking being a consequence of cell transfection is a report that studied the downregulating effect of a membrane-bound glycoprotein (GP) and described results obtained with its transfection in human embryonic kidney 293 (HEK293T) cells (Reynard *et al.*, 2009). In this paper, cells transfected with pCMLTrkA-GFP alone, showed a good correlation between the abundance of TrkA and GFP on flow cytometry. The authors, in effect, regarded GFP expression as a surrogate for TrkA presence and abundance. Next, in those cells cotransfected with pCMVGP and pCMLTrkA-GFP, about half of them while overexpressing GP and showing abundant GFP, revealed little TrkA expression on flow cytometry. Using confocal microscopy, the authors found that cells overexpressing GP revealed very low TrkA surface staining though high levels of TrkA-GFP remained

evident. Importantly, for those cells with moderate to no expression of GP, there was no effect on TrkA staining. The authors therefore concluded that epitope masking by overexpressed GP prevented the recognition of TrkA by its antibody. They however did not investigate the effect of protein knockdown to decide whether it was just the excessive amounts of protein present on overexpression or the non-specific effect of transfection that resulted in epitope masking. In agreement with the findings of Reynard *et al.* (2009), it could therefore be speculated that epitope masking was responsible for the lack of ER $\alpha$  detection on western blotting of transfected MCF7 cells noticed in the results described in chapter 5 of this thesis.

On account of the inability to demonstrate ER $\alpha$  protein expression changes after cells were transfected, QPCR assays were performed to determine whether changes in *CTCF* gene expression affected ER $\alpha$  gene expression. As noted previously, accurate and reproducible QPCR results are highly reliant on adequate optimisation of such processes as RNA extraction, primer structure, and sensitivity and specificity of the reaction via standard curves and melting points. In this section of the thesis, plasmid expression vectors were carefully studied via maxiprep and agarose gels electrophoresis. The plasmids were found to have a similar molecular weight to the expected. RNA extracted from transfected MCF7 cells was carefully assessed for quality and results fell well within acceptable levels as assessed by RIN using the Bioanalyser and the 260 / 280 ratio generated by spectrophotometry. All primers used in the experiments were tested to establish optimum performance with regards to their secondary structures. Using a bioinformatics tool, the primers were studied for secondary structures and found to have minimal folding confirming that there would be no primer interference with the annealing process. Careful pipetting was performed and the narrow difference between replicates (<0.8 Cq change) was evidence for replicate consistency. Linear range was determined to give guidance on the cDNA dilution range optimum for the experiments. The standard curves showed good efficiency for QPCR

doubling of template material as reactions were within the recommended range using CTCF and ER $\alpha$  primer pairs. Also of importance was the consistently single peak melting curves that showed a high degree of specificity for particular nucleic acid and primer pair combinations. Taken together, the results conformed to recommended MIQE guidelines and therefore gave a high degree of robustness to the data generated.

It is well recognised that there is often a disparity between protein expression and mRNA levels for many proteins as translation and post translational modifications could affect the total amount of protein generated for each mRNA (Gygi *et al.*, 1999; Newman *et al.*, 2006). In this thesis, an increase in mRNA expression for CTCF (table 5.3) was associated with a rise in CTCF protein expression (figure 5.5). Conversely, a fall in CTCF mRNA (Table 5.4) was reflected in low CTCF protein levels (figure 5.7B). This pattern of mRNA and protein expression levels is in agreement with Newman *et al.* (2006) who used DNA microarrays to detect changes in mRNA levels in yeast cells grown in different media. The authors noted that both mRNA and protein expression level changes occurred in the same direction and even in those cases in their report, in which mRNA expression did not tally with protein expression, further experiments with QPCR found that there had been errors in the DNA microarray data. Noteworthy in this thesis however, is the discrepancy in the extent of CTCF mRNA fold change relative to the protein expression change. As shown in figure 5.5, a ~70% CTCF protein overexpression was associated with a 16.08-fold mRNA upregulation (table 5.3). However, ~80% - 90% CTCF protein knockdown (figure 5.7) was apparently due to only a 4.54 fold decrease in CTCF mRNA expression (table 5.4). While it may seem surprising, this apparent discrepancy in the extent of mRNA knockdown relative to the associated protein expression change is not without precedent. Taiguchi *et al.* (2011) quantified mRNA levels with single molecule fluorescence *in situ* hybridization (smFISH) and RNA sequencing (RNA-seq) and constructed an imaging algorithm with a microfluidics device to determine single cell protein concentrations in *E. coli*. In

explaining the near-zero correlation between mRNA and protein expression levels in their results, they observed that mRNA and protein possess different life spans, with protein levels reflecting an accumulation of activity over cell cycles while mRNA levels reflecting minute to minute changes. This could explain the disparity in the extent of CTCF mRNA knockdown fold change relative to the protein expression knockdown.

The results obtained however indicated that there was no direct regulatory effect of CTCF mRNA expression on ER $\alpha$  gene expression as neither a 16.08-fold overexpression of CTCF mRNA nor a 4.54-fold knockdown of CTCF mRNA had any discernible effect on ER $\alpha$  mRNA expression levels. To the best of current knowledge, this is the first study that has attempted to determine a possible direct regulatory effect of *CTCF* gene expression change on *ER $\alpha$*  gene expression. Despite the previously described mutual interactions between CTCF and ER $\alpha$  (section 5.3), it is interesting, based on the results of chapter 5 in this thesis, that CTCF did not directly regulate ER $\alpha$  activity in the ER-positive MCF7 breast cancer cell line. This disparity could relate to the fact that the report of Ross-Ines *et al.* (2011) was based on bioinformatics and may not reflect biologically significant endogenous interactions. Also the report of Kawai *et al.* (2003) involved overexpression of HDAC1 that could result in nonbiological interactions as the protein in question is grossly exaggerated in the cell. Moreover, the latter authors did not examine the opposite effect (HDAC1 knockdown) to determine whether the converse was true for the interaction they found and they did not investigate a direct endogenous effect of CTCF on ER $\alpha$  expression.

Clearly, as shown in section 1.13 there is a complicated network of communication between CTCF and ER $\alpha$ . This network as suggested by the results described in this thesis including the mass spectrometry data analysed in Chapter 3 appeared to be indirect possibly mediated by EGFR signalling.



## CHAPTER 7

---

### Final discussion and conclusions

CTCF is a protein factor possessing 11 zinc fingers and capable of binding multiple DNA domains and in this way is able to exert an extensive array of intracellular effects. A 130kDa isoform had previously been discovered to be present in breast cancer tissue but not in normal breast tissue and a lower expression level of this isoform in patients with breast cancer was associated with worse prognostic indices (Docquier *et al.*, 2009). The mechanism(s) through which this CTCF isoform exerted its effects in breast cancer however is not known. There had previously been a suggestion that CTCF could be a proliferation factor based on its activity in primary cultures (Docquier *et al.*, 2009) therefore one of the aims of this thesis was to study the mechanistic involvement of CTCF in breast cancer via a possible association with other proliferation factors (Ki67 and PCNA) implicated to varying extents in the breast cancer phenotype. Furthermore, based on the predilection of CTCF for ER+ breast cancers, the thesis explored a possible direct regulatory relationship between CTCF and the estrogen receptor  $\alpha$ . Breast cancer cell lines possessing different hormone receptor (ER / PR) and HER2 profiles and different invasive potential and representing different biologic forms of the human disease were used to assess the mechanisms that governed CTCF action. The cells were investigated via indirect immunofluorescence to determine whether CTCF and the Ki67 / PCNA colocalised in the nuclei of breast cancer cell lines. Further experiments using immunoprecipitation, co-immunoprecipitation and western blotting were performed to assess physical binding between colocalising proteins. The findings from immunoprecipitation experiments were then confirmed with open label liquid chromatography / mass spectrometry (LC – MS / MS) assay. Based on previously published association between CTCF and estrogen receptor, and in combination with results from mass spectrometry in this thesis, further experiments were performed to

determine whether CTCF had a direct regulatory effect on ER $\alpha$  expression. Cells of an ER+ breast cancer cell line (MCF7) were transfected with a CTCF plasmid expression vector and CTCF siRNA to overexpress and knockdown CTCF expression respectively. Western blotting was performed to determine ER $\alpha$  expression in relation to variation in cellular CTCF protein expression. Furthermore, CTCF plasmid and siRNA transfected MCF7 cells were subjected to QPCR after reverse transcribing cDNA from extracted RNA. ER $\alpha$  gene expression change was then determined.

The results in this thesis revealed that CTCF-140 isoform, detected with a monoclonal antibody from BD Biosciences, was primarily expressed in the nucleolus of all the breast cancer cell lines studied. There was also some nucleoplasmic expression in the cell lines. Ki67 expression was mainly expressed in the nucleolus in all cell lines and PCNA expression though diffusely nucleoplasmic in all cell lines was also nucleoplasmic in the weakly invasive MCF7 and T47D cell lines. A clear nucleolar colocalisation between CTCF and Ki67 was observed in all breast cancer cell lines studied while colocalisation (also nucleolar) between CTCF and PCNA was seen in only the weakly invasive ER+ MCF7 and T47D cell lines. It is noteworthy that though all three proteins demonstrated nucleoplasmic expression, there was no colocalisation observed in this cellular compartment. The results pointed to a specific nucleolus-based activity for CTCF. Using western blotting and the monoclonal anti CTCF antibody from BD Biosciences, CTCF was found to be uniformly expressed in all cell lines studied though somewhat lower in the ER negative SKBR3 and MDA MB 231 cell lines. It was virtually not expressed in the normal luminal breast cell line LDM226. Indeed while Ki67 was expressed to a greater extent in the ER negative MDA MB 231 and LDM226 cell lines, PCNA expression was found to be less in these cell lines. It was not clear whether there was a reciprocal relationship between Ki67 and PCNA in these two cell lines. Though CTCF was successfully immunoprecipitated in all cell lines, there was no co-immunoprecipitation demonstrated with Ki67 and PCNA. Immunoprecipitated PCNA also demonstrated no

coprecipitation with CTCF or Ki67. In effect these experimental processes did not identify CTCF as being physically bound to either Ki67 or PCNA in these breast cancer cell lines. Furthermore and in contrast to previous reports, there was also no physical binding demonstrated between CTCF and both RNA pol II and PARP1 in breast cancer cell lines. Confirmation of these results was provided by open label liquid chromatography - mass spectrometry which further showed no physical binding between CTCF and ER / PR. It would appear therefore that the involvement of CTCF in breast cancer may not be explained via mechanisms through which Ki67 and PCNA mediated the breast cancer phenotype.

Careful inspection of the LC – MS / MS results however showed that CTCF bound with a high degree of certainty to three novel protein factors namely, general transcriptional factor 2 (GTF2), glucose related protein 78 (GRP78) and huntingtin interacting protein 1 related (HIP1r). A recently published work confirmed the CTCF–GTF2 interaction in a melanoma cell line. Examination of published work on these novel interacting proteins in tumorigenesis revealed that they had clear and extensive actions in epidermal growth factor receptor (EGFR) signalling. The association between CTCF and PI3K signalling (an EGFR signalling pathway) had previously been shown in a leukaemia cell line (Manavathi *et al.*, 2012) and lent strong support to the findings in this thesis. This thesis is probably the first to demonstrate an association in breast cancer cell lines between CTCF and EGFR signalling through novel interacting protein factors. The general transcriptional factor 2 is a component of the core transcriptional machinery in the cell and this physical binding demonstrated with CTCF suggested that the function of CTCF as a possible transcriptional factor could be through a protein partnership with GTF2. The results in this thesis are the first to demonstrate this association.

Further scrutiny of the mass spectrometry results showed that the novel CTCF protein partners were discovered in ER positive breast cancer cell lines. There is a known association between ER $\alpha$  and EGFR signalling and since mass spectrometry results

showed that CTCF was not physically bound to ER $\alpha$ , a possible direct regulatory effect of CTCF on ER $\alpha$  was therefore investigated. Overexpression and knockdown of CTCF in the hormone receptor positive MCF7 cells was associated with significant cell death confirming previous findings. Furthermore, both mRNA and protein expression moved in the same direction that is, a rise in mRNA was associated with an increase in protein expression and vice versa. Also, for some unclear reason, it appeared that CTCF transfection processes in MCF7 cells affected the ability to detect ER $\alpha$  protein expression on western blot assays. Importantly however, variation in CTCF mRNA expression did not reveal a change in ER $\alpha$  mRNA expression levels suggesting that CTCF did not directly regulate ER $\alpha$  mRNA expression in MCF7 cells. Again, this thesis could be the first to demonstrate this lack of direct regulatory effect of CTCF action on ER $\alpha$  expression.

The findings in this thesis would suggest that CTCF mediated the breast cancer phenotype specifically in ER positive cell lines not through an indirect effect at ER $\alpha$  signalling but via direct interaction with novel partners that have clear activities on EGFR signalling. The results would also suggest that exploitation of the interaction between CTCF and these novel protein interacting partners in ER positive breast cancer could extend the impact of the subclassification of breast cancer.

Much research and clinical trials have been performed trying to translate EGFR signalling interactions into druggable targets in breast cancer treatment but without much success (Masuda *et al.*, 2012). The results in this thesis provided a specific direction for further investigation regarding how CTCF mediated the breast cancer phenotype and provided another direction in the search for druggable targets in the EGFR pathway for breast cancer management. Further research would therefore involve interrogating the exact relationship between CTCF and the multiple intermediates involved in EGFR signalling.

## FUTURE DIRECTIONS

Cancer cell lines represent a model for experimental investigation and do not resemble the *de novo* cancer environment as the impact and influence of tissue stroma and an individual's innate immunity and psychological profile are lacking. This study needs to be performed on breast cancer tissue with different hormone receptor and HER2 phenotypes to confirm these novel CTCF protein partners.

The GTF2 – CTCF interaction with respect to EGFR signalling needs further investigation. There are multiple signalling pathways that stream from EGFR. It is needful to determine which particular path(s) the CTCF-GTF2 resides to further define that relationship.

## REFERENCES

- Abrantes, J. L., Tornatore, T. F., Pelizzaro-Rocha, K. J., de Jesus, M. B., Cartaxo, R. T., Milani, R., and Ferreira-Halder, C. V. Crosstalk between kinases, phosphatases and miRNAs in cancer. *Biochimie* <http://www.ncbi.nlm.nih.gov/pubmed/25230087> 2014; 107PB: 167-187.
- Acconcia, F. and Marino, M. The effects of 17beta-estradiol in cancer are mediated by estrogen receptor signalling at the plasma membrane. *Frontiers in Physiology* 2011; 2: 30.
- Alldrige, L., Metodieva, G., Greenwood, C., Al-Janabi, K., Thwaites, L., Sauven, P., and Metodiev, M. Proteome profiling of breast tumors by gel electrophoresis and nanoscale electrospray ionization mass spectrometry. *Journal of Proteome Research* 2008; 7(4): 1458-1469.
- Ames, H. M., Wang, A. A., Coughran, A., Evaul, K., Huang, S., Graves, C. W. and Ross, T. S. Huntingtin-Interacting protein 1 phosphorylation by receptor tyrosine kinases. *Molecular and Cellular Biology* 2013; 33 (18): 3580-3593.
- Anbalagan, M., Huderson, B., Murphy, L. and Rowan, B.G. Post-translational modifications of nuclear receptors and human disease. *Nuclear Receptor Signalling*, 2011; 10: e001-e001
- Ascenzi, P., Bocedi, A. and Marino, M. Structure–function relationship of estrogen receptor  $\alpha$  and  $\beta$ : impact on human health. *Molecular Aspects of Medicine* 2006; 27 (4): 299 – 402.
- Aulmann, S., Bläker, H., Penzel, R., Rieker, R.J., Otto, H.F. and Sinn, H.P. CTCF gene mutations in invasive ductal breast cancer. *Breast Cancer Research and Treatment* 2003; 80 (3): 347-352.

Bailey, T., Krajewski, P., Ladungs, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., and Zhang, J. Practical guidelines for the comprehensive analysis of ChIP-Seq data. *PLoS Computational Biology* 2013; 9 (11): e1003326.

Barrett, S.V. Breast cancer. *The Journal of the Royal College of Physicians of Edinburgh* 2010; 40 (4): 335-338.

Baydoun A. R. Cell culture techniques. In: Wilson, K. and Walker, J. eds. *Principles and Techniques of Biochemistry and Molecular Biology* 2010: 38 – 72. 7th ed. Cambridge: Cambridge University Press, UK.

Beneke S. Regulation of chromatin structure by poly(ADP-ribosyl)ation. *Frontiers in Genetics* 2012; 3: 169.

Bilal, E., Dutkowski, J., Guinney, J., Jang, I. S., Logsdon, B. A., Pandey, G. and Margolin, A. A. Improving breast cancer survival analysis through competition-based multidimensional modeling. *PLoS Computational Biology* 2013; 9 (5): e1003047.

Bogdanova, N. and Dörk, T. Molecular genetics of breast and ovarian cancer: recent advances and clinical implications. *Balkan Journal of Medical Genetics* 2012; 15 (Supl): 75-80.

Bolte, S., and Cordelieres, F. P. A guided tour into subcellular colocalisation analysis in light microscopy. *Journal of Microscopy* 2006; 224 (3): 213-232.

Brand, T.M., Lida, M., Luthar, N., Starr, M.M., Huppert, E.J. and Wheeler, D.L. Nuclear EGFR as a molecular target in cancer. *Radiotherapy and Oncology* 2013; 108 (3): 370-377.

Bridger, J.M., Kill, I.R. and Lichter, P. Association of pKi-67 with satellite DNA of the human genome in early G1 cells. *Chromosome Research* 1998; 6 (1), 13-24.

- Brooks, S.C., Locke, E.R. and Soule, H.D. Estrogen receptor in a human cell line (MCF-7) from breast carcinoma. *Journal of Biological Chemistry* 1973; 248 (17): 6251-6253.
- Bryant, H. E., Schultz, N., Thomas, H. D., Parker, K. M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, J. N., and Helleday, T. Specific killing of *BRCA2*-deficient tumours with inhibitors of poly (ADP-ribose) polymerase. *Nature* 2005; 434 (7035): 913-917.
- Bullwinkel, J., Baron-Lühr, B., Lüdemann, A., Wohlenberg, C., Gerdes, J. and Scholzen, T. Ki-67 protein is associated with ribosomal RNA transcription in quiescent and proliferating cells. *Journal of Cellular Physiology* 2006; 206 (3): 624-635.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. and Wittwer, C.T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 2009; 55 (4): 611 – 622.
- Butcher, D.T., Mancini-DiNardo, D., Archer, T.K. and Rodenhiser, D.I. DNA binding sites for putative methylation boundaries in the unmethylated region of the *BRCA1* promoter. *International Journal of Cancer* 2004; 111 (5): 669-678.
- Butcher, D.T. and Rodenhiser, D.I. Epigenetic inactivation of *BRCA1* is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. *European Journal of Cancer* 2007; 43 (1): 210-219.
- Cailleau R, Olive M, Cruciger QV. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro* 1978; 14 (11): 911 – 915.
- Carter, C.L., Allen, C. and Henson, D.E. Relation of tumour size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 1989; 63 (1): 181-187.



Celis, J.E. and Celis, A. Cell cycle-dependent variations in the distribution of the nuclear protein cyclin proliferating cell nuclear antigen in cultured cells: subdivision of S phase. *Proceedings of the National Academy of Sciences* 1985; 82 (10): 3262 - 3266.

Chan, C.S. and Song, J.S. CCCTC-binding factor confines the distal action of estrogen receptor. *Cancer Research* 2008; 68 (21): 9041 – 9049.

Chen, S., Iversen, E. S., Friebel, T., Finkelstein, D., Weber, B. L., Eisen, A., and Parmigiani, G. Characterization of BRCA1 and BRCA2 mutations in a large United States sample. *Journal of Clinical Oncology* 2006; 24 (6): 863-871.

Chen, S., and Parmigiani, G. Meta-analysis of BRCA1 and BRCA2 penetrance. *Journal of Clinical Oncology* 2007; 25 (11): 1329-1333.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., ... and Ng, H. H. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008; 133 (6): 1106 – 1117.

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergström, R., Kwon, Y., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., Morrison, I., Vigneron, M., Wu, S., Chiang, C., Loukinov, D., Lobanenko, V., Ohlsson, R. and Klenova, E. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Molecular and Cellular Biology* 2007; 27 (5): 1631-1648.

Choi, J. H., Min, N. Y., Park, J., Kim, J. H., Park, S. H., Ko, Y. J. and Lee, K. H. TSA-induced DNMT1 down-regulation represses hTERT expression via recruiting CTCF into demethylated core promoter region of hTERT in HCT116. *Biochemical and Biophysical Research Communications* 2010; 391 (1): 449-454.

Cicatiello, L., Addeo, R., Sasso, A., Altucci, L., Petrizzi, V. B., Borgo, R. and Weisz, A. Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. *Molecular and Cellular Biology* 2004; 24 (16): 7260-7274.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T., Cui, K. and Zhao, K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Research* 2009; 19 (1): 24 -32.

Dabbs, J.D. 2012. *Breast Pathology*. Philadelphia, USA: Elsevier, Saunders.

de Almeida, C. R., Stadhouders, R., de Bruijn, M. J., Bergen, I. M., Thongjuea, S., Lenhard, B., ... and Hendriks, R. W. The DNA-binding protein CTCF limits proximal V $\kappa$  recombination and restricts  $\kappa$  enhancer interactions to the immunoglobulin  $\kappa$  light chain locus. *Immunity* 2011; 35 (4), 501-513.

Debnath, J. and Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Reviews Cancer* 2005; 5 (9): 675 – 688.

Defosse, P., Kelly, K.F., Filion, G.J.P., Pérez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C.L., Daniel, J.M. and Gilson, E. The human enhancer blocker CCCTC-binding factor interacts with the transcription factor Kaiso. *Journal of Biological Chemistry* 2005; 280 (52): 43017-43023.

Del Campo, E.P., Talamás Márquez, J.J., Reyes-Vargas, F., Del Pilar Intriago-Ortega, M., Quintanar-Escorza, M.A., Burciaga-Nava, J.A., Sifuentes-Alvarez, A. and Reyes-Romero, M. CTCF and CTCFL mRNA expression in 17 $\beta$ -estradiol-treated MCF7 cells. *Biomedical Report* 2014; 2 (1): 101-104.

De Smet, C., Lorient, A., and Boon, T. Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells. *Molecular and Cellular Biology* 2004; 24 (11): 4781-4790.

Docquier, F., Farrar, D., D'Arcy, V., Chernukhin, I., Robinson, A.F., Loukinov, D., Vatolin, S., Pack, S., Mackay, A., Harris, R.A., Dorricott, H., O'Hare, M., J., Lobanenko, V. and Klenova, E. Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Research* 2005; 65 (12): 5112-5122.

Docquier, F., Kita, G., Farrar, D., Jat, P., O'Hare, M., Chernukhin, I., Gretton, S., Mandal, A., Alldridge, L. and Klenova, E. Decreased poly(ADP-ribosyl)ation of CTCF, a transcription factor, is associated with breast cancer phenotype and cell proliferation. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 2009; 15 (18): 5762-5771.

Domchek, S. M., Friebel, T. M., Singer, C. F., Evans, D. G., Lynch, H. T., Isaacs, C., Garber J. E., Neuhausen S. L., Matloff, E., Eeles, R., Pichert, G., Van t'Veer, L., Tung, N., Weitzel, J. N., Couch, F. L., Rubinstein, S. W., Ganz, P. A., Daly, M. B., Olopade, I. O., Tomlinson, G., Schildkraut, J., Blum, J. L. and Rebbeck, T. R. Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *Journal of the American Medical Association* 2010; 304 (9): 967-975.

Dong, D., Ni, M., Li, J., Xiong, S., Ye, W., Virrey, J.J., Mao, C., Ye, R., Wang, M., Pen, L., Dubeau, L., Groshen, S., Hofman, F.M. and Lee, A.S. Critical role of the stress chaperone GRP78 / BiP in tumour proliferation, survival, and tumour angiogenesis in transgene-induced mammary tumour development. *Cancer Research* 2008; 68 (2): 498-505.

Donohoe, M.E., Zhang, L., Xu, N., Shi, Y. and Lee, J.T. Identification of a CTCF cofactor, Yy1, for the X chromosome binary switch. *Molecular Cell* 2007; 25 (1): 43-56.

Dowsett, M., Nielsen, T.O., A'Hern, R., Bartlett, J., Coombes, R.C., Cuzick, J., Ellis, M., Henry, N.L., Hugh, J.C. and Lively, T. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *Journal of the National Cancer Institute* 2011; 103 (22): 1656 -1664.

Dundr, M., Misteli, T. and Olson, M.O. The dynamics of post mitotic reassembly of the nucleolus. *Journal of Cell Biology* 2000; 150 (3): 433-446.

Eilers, M. and Eisenman, R, N. *Myc's* broad reach. *Genes and Development* 2008; 22 (20): 2755-2766.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M. D., Zimonjic, D. B., Donaher, J. L. and Weinberg, R. A. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes and Development* 2001; 15 (1): 50-65.

El-Kady, A. and Klenova, E. Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Letters* 2005; 579 (6): 1424-1434.

Engel, N., Thorvaldsen, J.L. and Bartolomei, M.S. CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19 / Igf2 locus. *Human Molecular Genetics* 2006; 15 (19): 2945-2954.

Engelken, F., Bremme, R., Bick, U., Hammann-Kloss, S. and Fallenberg, E.M. Factors affecting the rate of false positive marks in CAD in full-field digital mammography. *European Journal of Radiology* 2012; 81 (8): e844 - e848.

Essien, K., Vigneau, S., Apreleva, S., Singh, L.N., Bartolomei, M.S. and Hannenhalli, S. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biology* 2009; 10 (11): R131-R131.

Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics* 2007; 8 (4): 286-298.

Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A. and Klenova, E. Mutational analysis of the poly(ADP-ribosylation) sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosylation). *Molecular and Cellular Biology* 2010; 30 (5): 1199-1216.

Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., and Bartolomei, M.S. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* 2004; 303 (5655): 238-240.

Fedotov, S., Iomin, A. and Ryashko, L. Non-Markovian models for migration-proliferation dichotomy of cancer cells: Anomalous switching and spreading rate. *Physical Review E* 2011; 84 (6): 061131.

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G. and Lobanenkov, V. V. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology* 1996; 16 (6): 2802-2813.

Fischer, E. Antibody Engineering in Translational Medicine. *Engineering in Translational Medicine*. Springer; 2014: 301-316.

Fitzgibbons, P.L., Page, D.L., Weaver, D., Thor, A.D., Allred, D.C., Clark, G.M., Ruby, S.G., O'Malley, F., Simpson, J.F., Connolly, J.L., Hayes, D.F., Edge, S.B., Lichter, A.

and Schnitt, S.J. Prognostic Factors in Breast Cancer. *Archives of Pathology and Laboratory Medicine* 2000; 124 (7): 966-978.

Fleige, S. and Pfaffl, M.W. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* 2006; 27 (2): 126 - 139.

Ford, C.H., Al-Bader, M., Al-Ayadhi, B. and Francis, I. Reassessment of estrogen receptor expression in human breast cancer cell lines. *Anticancer Research* 2011; 31 (2): 521-527.

Fowler, C. B., Evers, D. L., O'Leary, T. J., and Mason, J. T. Antigen retrieval causes protein unfolding evidence for a linear epitope model of recovered immunoreactivity. *Journal of Histochemistry and Cytochemistry* 2011; 59 (4): 366-381.

Fox, J.T., Lee, K. and Myung, K. Dynamic regulation of PCNA ubiquitylation / deubiquitylation. *FEBS Letters* 2011; 585 (18): 2780-2785.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A. and Jensen, L. J. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research* 2013; 41 (D1): D808 - D815.

Frouin, I., Maga, G., Denegri, M., Riva, F., Savio, M., Spadari, S., Prosperi, E. and Scovassi, A.I., Human proliferating cell nuclear antigen, poly(ADP-ribose) polymerase-1, and p21waf1/cip1. A dynamic exchange of partners. *Journal of Biological Chemistry* 2003; 278 (41): 39265-39268.

Fuhrman, B. J., Schairer, C., Gail, M. H., Boyd-Morin, J., Xu, X., Sue, L. Y. and Ziegler, R. G. Estrogen metabolism and risk of breast cancer in postmenopausal women. *Journal of the National Cancer Institute* 2012; 104 (4): 326 - 339.

Gerdes, J., Li, L., Schlueter, C., Duchrow, M., Wohlenberg, C., Gerlach, C., Stahmer, I., Kloth, S., Brandt, E. and Flad, H. Immunobiochemical and molecular biologic

characterization of the cell proliferation-associated nuclear antigen that is defined by monoclonal antibody Ki-67. *American Journal of Pathology* 1991; 138 (4): 867.

Geyer C.F., Reis-Filho J. S. and Dabbs J. D. Lobular neoplasia and invasive lobular carcinoma. In: Dabbs, D., ed. *Breast Pathology* 2012. Philadelphia, USA: Elsevier, Saunders. 1 - 21.

Gill, G. Something about SUMO inhibits transcription. *Current Opinion in Genetics and Development* 2005; 15 (5): 536-541.

Gingras, A., Gstaiger, M., Raught, B. and Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nature Reviews: Molecular Cell Biology* 2007; 8 (8): 645-654.

Girardot, M., Cavaillé, J., and Feil, R. Small regulatory RNAs controlled by genomic imprinting and their contribution to human disease. *Epigenetics* 2012; 7 (12): 1341-1348.

Gómez, D. L. M., Farina, H. G., and Gómez, D. E. Telomerase regulation: A key to inhibition? *International Journal of Oncology* 2013; 43 (5): 1351 - 1356.

Gordon, S., Akopyan, G., Garban, H. and Bonavida, B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* 2006; 25 (8): 1125-1142.

Green, A. R., Krivinskas, S., Young, P., Rakha, E. A., Paish, E. C., Powe, D. G., and Ellis, I. O. Loss of expression of chromosome 16q genes *DPEP1* and *CTCF* in lobular carcinoma in situ of the breast. *Breast Cancer Research and Treatment* 2009; 113 (1): 59-66.

Gygi, S. P., Rochon, Y., Franza, B. R., & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology* 1999; 19 (3): 1720-1730.

Hammes, S.R. and Levin, E.R. Extranuclear steroid receptors: nature and actions. *Endocrine Reviews* 2007; 28 (7): 726-741.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M. and Wei, C. L. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics* 2011; 43 (7): 630 - 638.

Hannenhalli, S. and Wang, L. Enhanced position weight matrices using mixture models. *Bioinformatics* 2005; 21 (Suppl 1): i204-i212.

Heath, H., Ribeiro, D.A., Sleutels, F., Dingjan, G., van, d.N., Jonkers, I., Ling, K., Gribnau, J., Renkawitz, R., Grosveld, F., Hendriks, R.W. and Galjart, N. CTCF regulates cell cycle progression of alpha beta T cells in the thymus. *European Molecular Biology Organisation Journal* 2008; 27 (21): 2839-2850.

Hendrickson, E. L., Xia, Q., Wang, T., Leigh, J. A., and Hackett, M. Comparison of spectral counting and metabolic stable isotope labelling for use with quantitative microbial proteomics. *Analyst* 2006; 131(12): 1335-1341.

Hervouet, E., Cartron, P.F., Jouvenot, M. and Delage-Mourroux, R. Epigenetic regulation of estrogen signalling in breast cancer. *Epigenetics: Official Journal of the DNA Methylation Society* 2013; 8 (3): 237-245.

Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T. CTCF is required for neural development and stochastic expression of clustered *Pcdh* genes in neurons. *Cell Reports* 2012; 2: 345–357.

Hoda, S. Normal breast and developmental disorders. In: Dabbs, D., ed. *Breast Pathology* 2012. Philadelphia, USA: Elsevier, Saunders.1 – 21.

Holliday, D., and Speirs, V. Choosing the right cell line for breast cancer research. *Breast Cancer Research* 2011; 13(4): 215



Hou, M., Kuo, H., Li, J., Wang, Y., Chang, C., Chen, K., Chen, W., Chiu, C., Yang, S. and Chang, W. Orai1/CRACM1 overexpression suppresses cell proliferation via attenuation of the store-operated calcium influx-mediated signalling pathway in A549 lung cancer cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2011; 1810 (12): 1278-1284.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D. and Carroll, J.S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature Genetics* 2011; 43 (1): 27-33.

Hyun, T. S., Rao, D. S., Saint-Dic, D., Michael, L. E., Kumar, P. D., Bradley, S. V. And Ross, T. S. HIP1 and HIP1r stabilize receptor tyrosine kinases and bind 3-phosphoinositides via epsin N-terminal homology domains. *Journal of Biological Chemistry* 2004; 279 (14): 14294 -14306.

Ishii K., Laemmli U.K. Structural and dynamic functions establish chromatin domains. *Molecular Cell* 2003; 11: 237 – 248.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. Global cancer statistics. *CA: A Cancer Journal for Clinicians* 2011; 61 (2): 69-90.

Jesinger, R. A. Breast anatomy for the interventionalist. *Techniques in Vascular and Interventional Radiology* 2014; 17 (1): 3-9.

Jothi, R., Cuddapah, S., Barski, A., Cui, K. and Zhao, K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Research* 2008; 36 (16): 5221-5231.

Jovanovic, J., Rønneberg, J. A., Tost, J., and Kristensen, V. The epigenetics of breast cancer. *Molecular Oncology* 2010; 4 (3): 242-254.

Kaake, R.M., Wang, X. and Huang, L. Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Molecular and Cellular Proteomics* 2010; 9 (8): 1650-1665.

Kawai, H., Li, H., Avraham, S., Jiang, S. and Avraham, H.K. Overexpression of histone deacetylase HDAC1 modulates breast cancer progression by negative regulation of estrogen receptor. *International Journal of Cancer* 2003; 107 (3): 353-358.

Khidr, L. and Chen, P. RB, the conductor that orchestrates life, death and differentiation. *Oncogene* 2006; 25: 5210 – 5219.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenko, V.V. and Ren, B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007; 128 (6): 1231-1245.

Kitchen, N. S. and Schoenherr, C. J. Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin. *Journal of Cellular Biochemistry* 2010; 111 (3): 665-675.

Klajic, J., Busato, F., Edvardsen, H., Touleimat, N., Fleischer, T., Bukholm, I. R., and Kristensen, V. N. DNA Methylation status of key cell cycle regulators such as CDKN2A/p16 and CCNA1 correlates with treatment response to doxorubicin and 5-fluorouracil in locally advanced breast tumours. *Clinical Cancer Research* 2014; clincanres-0297.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E. and Lobanenko, V.V. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Molecular and Cellular Biology* 1993; 13 (12): 7612-7624.

- Klenova, E. M., Chernukhin, I. V., El-Kady, A., Lee, R. E., Pugacheva, E. M., Loukinov, D. I. and Lobanenko, V. V. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Molecular and Cellular Biology* 2001; 21 (6): 2221 - 2234.
- Klenova, E., Chernukhin, I., Inoue, T., Shamsuddin, S., and Norton, J. Immunoprecipitation techniques for the analysis of transcription factor complexes. *Methods* 2002; 26 (3): 254-259.
- Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenko, V., Reik, W. and Ohlsson, R. CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proceedings of the National Academy of Sciences USA* 2006; 103, 10684-10689.
- La Rosa-Vela'zquez, de I. A., Rinco'n-Arango, H., Benítez-Bribiesca, L., and Recillas-Targa, F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Research* 2007; 67: 2577 – 2585.
- Lacroix, M., and Leclercq, G. Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Research and Treatment* 2004; 83(3): 249-289.
- Laity, J. H., Lee, B. M., and Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology* 2001; 11(1), 39-46.
- Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. and Cremer, T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews Genetics* 2007; 8 (2): 104-115.

Lasfargues, E. Y., Coutinho, W. G., and Dion, A. S. A human breast tumour cell line (BT-474) that supports mouse mammary tumour virus replication. *In vitro* 1979; 15 (9): 723 - 729.

Lauand, C., Rezende-Teixeira, P., Cortez, B.A., de Oliveira Niero, Evandro Luís and Machado-Santelli, G.M. Independent of ErbB1 gene copy number, EGF stimulates migration but is not associated with cell proliferation in non-small cell lung cancer. *Cancer Cell International* 2013; 13 (1): 38

Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, Macbeath G. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signalling networks. *Cell* 2012; 149: 780 – 794.

Leung, A.K. and Lamond, A.I. The dynamics of the nucleolus. *Critical Reviews™ in Eukaryotic Gene Expression* 2003;13 (1): 39-54.

Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* 2002; 3 (9): 662-673.

Li, T. and Lu, L. Epidermal growth factor-induced proliferation requires down-regulation of Pax6 in corneal epithelial cells. *Journal of Biological Chemistry* 2005; 280 (13): 12988 - 12995.

Liao, X., Lu, D., Wang, N., Liu, L., Wang, Y., Li, Y., Yan, T., Sun, X., Hu, P. and Zhang, T. Estrogen receptor  $\alpha$  mediates proliferation of breast cancer MCF-7 cells via a p21/PCNA/E2F1-dependent pathway. *FEBS Journal* 2014; 281 (3): 927 - 42.

Ling, J.Q., Li, T., Hu, J.F., Vu, T.H., Chen, H.L., Qiu, X.W., Cherry, A.M. and Hoffman, A.R. CTCF mediates interchromosomal colocalisation between Igf2 / H19 and Wsb1 / Nf1. *Science* 2006; 312 (5771): 269-272.

Liu, S. and Lin, Y.C. Transformation of MCF-10A human breast epithelial cells by zeranol and estradiol-17 $\beta$ . *Breast Journal* 2004; 10 (6): 514-521.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., and Goodwin, G. H. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 1990; 5 (12): 1743-1753.

Lundgren, D. H., Hwang, S. I., Wu, L., and Han, D. K. Role of spectral counting in quantitative proteomics. *Expert Review of Proteomics* 2010; 7 (1): 39-53.

Luo, B. and Lee, A.S. The critical roles of endoplasmic reticulum chaperones and unfolded protein response in tumorigenesis and anticancer therapies. *Oncogene* 2013; 32 (7): 805-818.

Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V. and Renkawitz, R. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Research* 2000; 28 (8): 1707-1713.

MacCallum, D.E. and Hall, P.A. The location of pKi67 in the outer dense fibrillary compartment of the nucleolus points to a role in ribosome biogenesis during the cell division cycle. *Journal of Pathology* 2000; 190 (5): 537-544.

Maciejczyk, A. New prognostic factors in breast cancer. *Advances in Clinical and Experimental Medicine* 2013; 22: 5-15.

MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M., and Sadowski, P. D. The CTCF insulator protein is post translationally modified by SUMO. *Molecular and Cellular Biology* 2009; 29 (3): 714-725.

Majka, J. and Burgers, P.M.J. The PCNA – RFC families of DNA clamps and clamp loaders. *Progress in Nucleic acid Research and Molecular Biology* 2004; 78: 227-260.

Majumder, P., Gomez, J.A., Chadwick, B.P. and Boss, J.M. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *Journal of Experimental Medicine* 2008; 205 (4): 785-798.

Majumder, P., and Boss, J., M. CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. *Molecular and Cell Biology* 2010; 30: 4211– 4223.

Manavathi, B., Lo, D., Bugide, S., Dey, O., Imren, S., Weiss, M. J., and Humphries, R. K. Functional regulation of pre-B-cell leukemia homeobox interacting protein 1 (PBXIP1 / HPIP) in erythroid differentiation. *Journal of Biological Chemistry* 2012; 287 (8): 5600 - 5614.

Marchio, C., Natrajan, R., Shiu, K., Lambros, M., Rodriguez-Pinilla, S., Tan, D., Lord, C., Hungermann, D., Fenwick, K. and Tamber, N. The genomic profile of HER2-amplified breast cancers: the influence of ER status. *Journal of Pathology* 2008; 216 (4): 399-407.

Maringe C, Mangtani P, Rachet B, Leon DA, Coleman MP, dos Santos Silva I. Cancer incidence in South Asian migrants to England, 1986–2004: Unraveling ethnic from socioeconomic differentials. *International Journal of Cancer* 2013; 132 (8): 1886-94.

Marino, M., and Ascenzi, P. Membrane association of estrogen receptor  $\alpha$  and  $\beta$  influences 17 $\beta$ -estradiol-mediated cancer cell proliferation. *Steroids* 2008; 73 (9): 853-858.

Markham, N.R. and Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research* 2005; 33 (suppl 2): 577- 581.

- Martinez, S.R. and Miranda, J.L. CTCF terminal segments are unstructured. *Protein Science* 2010; 19 (5): 1110-1116.
- Mason, M., Schuller, A., and Skordalakes, E. Telomerase structure function. *Current Opinion in Structural Biology* 2011; 21 (1): 92-100.
- Masuda, H., Zhang, D., Bartholomeusz, C., Doihara, H., Hortobagyi, G. N., and Ueno, N. T. Role of epidermal growth factor receptor in breast cancer. *Breast Cancer Research and Treatment* 2012; 136 (2): 331-345.
- Masutomi, K., Yu, E. Y., Khurts, S., Ben-Porath, I., Currier, J. L., Metz, G. B. and Hahn, W. C. Telomerase maintains telomere structure in normal human cells. *Cell* 2003; 114 (2): 241-253.
- Matros, E., Wang, Z.C., Lodeiro, G., Miron, A., Iglehart, J.D. and Richardson, A.L. *BRCA1* promoter methylation in sporadic breast tumours: relationship to gene expression profiles. *Breast Cancer Research and Treatment* 2005; 91 (2): 179-186.
- Meeran, S. M., Patel, S. N., Tollefsbol T. O. Sulforaphane causes epigenetic repression of hTERT expression in human breast cancer cell lines. *PLoS ONE* 2010; 5: e11457.
- Melchor, L., Molyneux, G., Mackay, A., Magnay, F., Atienza, M., Kendrick, H., Nava-Rodrigues, D., López-García, M.Á., Milanezi, F., Greenow, K., Robertson, D., Palacios, J., Reis-Filho, J.S. and Smalley, M.J. Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models. *Journal of Pathology* 2014; 233 (2): 124-137.
- Mendez, M. G., Kojima, S. I., and Goldman, R. D. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *Federation of American Societies for Experimental Biology Journal* 2010; 24 (6): 1838-1851.

Miller, E., Lee, H. J., Lulla, A., Hernandez, L., Gokare, P. and Lim, B. Current treatment of early breast cancer: adjuvant and neoadjuvant therapy. *F1000Research* 2014; 3: 198.

Minning C., Mokhtar N.M., Abdullah N., Muhammad R., Emran N.A., Ali M.A.S., Harun R., and Jamal R. Exploring breast carcinogenesis through integrative genomics and epigenomics analyses. *International Journal of Oncology* 2014; 45 (5): 1959 - 68.

Misra, U., Wang, F. and Pizzo, S. Transcription factor TFII-I causes transcriptional upregulation of GRP78 synthesis in prostate cancer cells. *Journal of Cellular Biochemistry* 2009; 106 (3): 381-389.

Moldovan, G.L., Pfander, B. and Jentsch, S. PCNA, the maestro of the replication fork. *Cell* 2007; 129 (4): 665-679.

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S. T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., Burke, L.J., Renkawitz-Pohl, R., Ohlsson, R., Zhou, J., Renkawitz, R. and Lobanenko, V. CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *European Molecular Biology Organisation Reports* 2005; 6 (2): 165-170.

Moran, M. S., Schnitt, S. J., Giuliano, A. E., Harris, J. R., Khan, S. A., Horton, J., Klimberg S, Chavez-MacGregor, M., Freedman, G., Houssami, N., Johnson, P. L. and Morrow, M. Society of Surgical Oncology–American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer. *Journal of Clinical Oncology* 2014; 32 (14): 1507-1515.

Murphy, L. C., Seekallu, S. V., and Watson, P. H. Clinical significance of estrogen receptor phosphorylation. *Endocrine-Related Cancer* 2011; 18 (1): R1-R14.



Murrell, A., Heeson, S., and Reik, W. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nature Genetics* 2004; 36: 889 – 893.

Musinova, Y.R., Lisitsyna, O.M., Golyshev, S.A., Tuzhikov, A.I., Polyakov, V.Y. and Sheval, E.V. Nucleolar localization/retention signal is responsible for transient accumulation of histone H2B in the nucleolus through electrostatic interactions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2011; 1813 (1): 27-38.

Nasu, K., Takai, N., Nishida, M., and Narahara, H. Tumorigenic effects of tamoxifen on the female genital tract. *Clinical Medicine. Pathology* 2008; 1: 17 - 34.

Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L. and Weissman, J.S. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 2006; 441 (7095): 840-846.

Nilsson, M. P., Hartman, L., Idvall, I., Kristoffersson, U., Johannsson, O. T., and Loman, N. Long-term prognosis of early-onset breast cancer in a population-based cohort with a known BRCA1/2 mutation status. *Breast Cancer Research and Treatment* 2014; 144 (1): 133-142.

Ohlsson, R., Renkawitz, R., and Lobanenko, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics* 2001; 17(9): 520-527.

Ohlsson, R., Lobanenko, V. and Klenova, E. Does CTCF mediate between nuclear organization and gene expression? *Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology* 2010; 32 (1): 37-50.

O'Hare, M. J., Bond, J., Clarke, C., Takeuchi, Y., Atherton, A. J., Berry, C. and Jat, P. S. Conditional immortalization of freshly isolated human mammary fibroblasts and

endothelial cells. *Proceedings of the National Academy of Sciences* 2001; 98 (2): 646-651.

O'Malley, B.W. and Kumar, R. Nuclear receptor coregulators in cancer biology. *Cancer Research* 2009; 69 (21), 8217 - 8222.

Ong, C. T., and Corces, V. G. Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme. *Journal of Biology* 2009; 8: 73.

Osaki, L., Figueiredo, P., Alvares, E. and Gama, P. EGFR is involved in control of gastric cell proliferation through activation of MAPK and Src signalling pathways in early-weaned rats. *Cell Proliferation* 2011; 44 (2): 174-182.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. and Fraser, P. Active genes dynamically colocalise to shared sites of ongoing transcription. *Nature Genetics* 2004; 36 (10): 1065-1071.

Ouelle, D. E., Zindy, F., Ashmun, R. A., and Sherr, C. J. Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell* 1995; 83 (6): 993 - 1000.

Ozenne, P., Eymin, B., Brambilla, E., and Gazzeri, S., The ARF tumor suppressor: Structure, functions and status in cancer. *International Journal of Cancer* 2010; 127: 2239 – 2247.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G. and Merkenschlager, M. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 2008; 132 (3): 422-433.

- Park, S., Koo, J. S., Kim, M. S., Park, H. S., Lee, J. S., Lee, J. S. and Park, B. W. Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry. *Breast* 2012; 21(1): 50-57.
- Patani, N., Martin, L. and Dowsett, M. Biomarkers for the clinical management of breast cancer: international perspective. *International Journal of Cancer* 2013; 133 (1): 1 – 13.
- Paul, F.E., Hosp, F. and Selbach, M. Analyzing protein–protein interactions by quantitative mass spectrometry. *Methods* 2011; 54 (4): 387-395.
- Payer, B. and Lee, J.T. X chromosome dosage compensation: how mammals keep the balance. *Annual Review of Genetics* 2008; 42: 733-772.
- Peña-Hernández, R., Marques, M., Hilmi, K., Zhao, T., Saad, A., Alaoui-Jamali, M. A., del Rincon, S.V., Ashworth, T., Ananda, L. R., Emerson, B. M., and Witcher, M. Genome-wide targeting of the epigenetic regulatory protein CTCF to gene promoters by the transcription factor TFII-I. *Proceedings of the National Academy of Sciences* 2015; 112 (7): E677-E686.
- Phillips, J.E. and Corces, V.G. CTCF: master weaver of the genome. *Cell* 2009; 137 (7): 1194-1211.
- Piccart-Gebhart, M.J., Procter, M., Leyland-Jones, B., Goldhirsch, A., Untch, M., Smith, I., Gianni, L., Baselga, J., Bell, R. and Jackisch, C. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *New England Journal of Medicine* 2005; 353 (16): 1659-1672.
- Pocobelli, G., and Weiss, N. S. Breast cancer mortality in relation to receipt of screening mammography: a case–control study in Saskatchewan, Canada. *Cancer Causes and Control* 2014; 1-7.

Prat, A., and Baselga, J. The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of HER2. *Nature Clinical Practice Oncology* 2008; 5 (9): 531 – 542.

Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., and Perou, C. M. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research* 2010; 12 (5): R68.

Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., Quitschke, W.W., Loukinov, D.I., Ohlsson, R. and Lobanenko, V.V. Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Human Molecular Genetics* 2005; 14 (7): 953 - 965.

Qi, C. F., Martensson, A., Mattioli, M., Dalla-Favera, R., Lobanenko, V. V., and Morse, H. C. CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. *Proceedings of the National Academy of Sciences* 2003; 100 (2): 633 - 638.

Qiu, C., Shan, L., Yu, M., and Snyderwine, E. G. Steroid hormone receptor expression and proliferation in rat mammary gland carcinomas induced by 2-amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine. *Carcinogenesis* 2005; 26 (4): 763 - 769.

Rahmanzadeh, R., Hüttmann, G., Gerdes, J. and Scholzen, T. Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis. *Cell Proliferation* 2007; 40 (3): 422-430.

Rakha, E.A., Pinder, S.E., Paish, C.E. and Ellis, I.O. Expression of the transcription factor CTCF in invasive breast cancer: a candidate gene located at 16q22.1. *British Journal of Cancer* 2004; 91 (8): 1591-1596.

- Rakha, E.A., Armour, J.A.L., Pinder, S.E., Paish, C.E. and Ellis, I.O. High-resolution analysis of 16q22.1 in breast carcinoma using DNA amplifiable probes (multiplex amplifiable probe hybridization technique) and immunohistochemistry. *International Journal of Cancer* 2005; 114 (5): 720-729.
- Rakha E, A., Green A. R., Powe D. G., Roylance R., Ellis I. O. Chromosome 16 tumour-suppressor genes in breast cancer. *Genes, Chromosomes and Cancer* 2006; 45: 527–535.
- Rakha E. A. and Ellis I. O. Ductal carcinoma in situ. In: Dabbs, D., ed. *Breast Pathology* 2012. Philadelphia, USA: Elsevier, Saunders. 1 - 21.
- Ramaswamy B, Fiskus W, Cohen B, Pellegrino C, Hershman DL, Chuang E. Phase I-II study of vorinostat plus paclitaxel and bevacizumab in metastatic breast cancer: Evidence for vorinostat-induced tubulin acetylation and Hsp90 inhibition *in vivo*. *Breast Cancer Research and Treatment* 2012; 132: 1063 – 72.
- Ramos, E. A., Camargo, A. A., Braun, K., Slowik, R., Cavalli, I. J., Ribeiro, E. M., ... and Klassen, G. Simultaneous CXCL12 and ESR1 CpG island hypermethylation correlates with poor prognosis in sporadic breast cancer. *BMC Cancer* 2010; 10 (1): 23
- Rasko, J.E., Klenova, E.M., Leon, J., Filippova, G.N., Loukinov, D.I., Vatolin, S., Robinson, A.F., Hu, Y.J., Ulmer, J., Ward, M.D., Pugacheva, E.M., Neiman, P.E., Morse, H.C., 3, Collins, S.J. and Lobanenkov, V.V. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Research* 2001; 61 (16): 6002 - 6007.
- Razin, S. V., Borunova, V. V., Maksimenko, O. G., and Kantidze, O. L. Cys2His2 zinc finger protein family: classification, functions, and major members. *Biochemistry (Moscow)* 2012; 77 (3): 217-226.

Rebbeck, T. R., Friebel, T. and Lynch, H. T. Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers. The PROSE Study Group. *Journal of Clinical Oncology* 2004; 22: 1055 – 1062.

Recillas-Targa, F., Rosa-Velázquez, I. A., Soto-Reyes, E., and Benítez-Bribiesca, L. Epigenetic boundaries of tumour suppressor gene promoters: the CTCF connection and its role in carcinogenesis. *Journal of Cellular and Molecular Medicine* 2006; 10 (3): 554-568.

Reik, W., Murrell, A., Lewis, A., Mitsuya, K., Umlauf, D., Dean, W., Higgins, M. and Feil, R. Chromosome loops, insulators, and histone methylation: new insights into regulation of imprinting in clusters. *Cold Spring Harbor Symposia on Quantitative Biology* 2004; 69: 29-37.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G. and Pedone, P.V. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *Journal of Biological Chemistry* 2007; 282 (46): 33336-33345.

Reynard, O., Borowiak, M., Volchkova, V.A., Delpeut, S., Mateo, M. and Volchkov, V.E. Ebolavirus glycoprotein GP masks both its own epitopes and the presence of cellular surface proteins. *Journal of Virology* 2009; 83 (18): 9596 - 9601.

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M. J., Bergen, I. M., Thongjuea, S., Lenhard, B. and Hendriks, R. W. The DNA-binding protein CTCF limits proximal V $\kappa$  recombination and restricts  $\kappa$  enhancer interactions to the immunoglobulin  $\kappa$  light chain locus. *Immunity* 2011; 35 (4): 501-513.

Rodriguez, C., Borgel, J., Court, F., Cathala, G., Forne', T., and Piette, J. CTCF is a DNA methylation-sensitive positive regulator of the *INK/ARF* locus. *Biochemical and Biophysical Research Communications* 2010; 392: 129–134.

Rodriguez, B.A., Weng, Y.I., Liu, T.M., Zuo, T., Hsu, P.Y., Lin, C.H., Cheng, A.L., Cui, H., Yan, P.S. and Huang, T.H. Estrogen-mediated epigenetic repression of the imprinted gene cyclin-dependent kinase inhibitor 1C in breast cancer cells. *Carcinogenesis* 2011; 32 (6): 812-821.

Rosen, P. P., Groshen, S., Kinne, D., and Norton, L. Factors influencing prognosis in node-negative breast carcinoma: analysis of 767 T1N0M0 / T2N0M0 patients with long-term follow-up. *Journal of Clinical Oncology* 1993; 11 (11): 2090-2100.

Ross-Innes, C.S., Brown, G.D. and Carroll, J.S. A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC Genomics* 2011; 12: 593-2164-12-593.

Roy, A.L. Signal-induced functions of the transcription factor TFII-I. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 2007; 1769 (11–12): 613-621.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteché, C. M., Filippova, G. N., Baliga, N. S. and Krumm, A. CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences* 2008; 105 (24): 8309 - 8314.

Russo, J. and Russo, I.H. The role of estrogen in the initiation of breast cancer. *Journal of Steroid Biochemistry and Molecular Biology* 2006; 102 (1–5): 89-96.

Sabnis, G.J., Goloubeva, O., Chumsri, S., Nguyen, N., Sukumar, S. and Brodie, A.M. Functional activation of the estrogen receptor-alpha and aromatase by the HDAC inhibitor entinostat sensitizes ER-negative tumors to letrozole. *Cancer Research* 2011; 71 (5): 1893 - 1903.

Savouret, J. F., Fridlanski, F., Atger, M., Misrahi, M., Berger, R., and Milgrom, E. Origin of the high constitutive level of progesterone receptor in T47D breast cancer cells. *Molecular and Cellular Endocrinology* 1991; 75 (2): 157 - 162.

Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., and Jasin, M. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* 2011; 145 (4): 529-542.

Schlüter, C. The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins. *Journal of Cell Biology* 1993; 123 (3): 513.

Scholzen, T. and Gerdes, J. The Ki-67 protein: from the known and the unknown. *Journal of Cellular Physiology* 2000; 182 (3): 311-322.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 2006; 7: 3

Segura-Puimedon, M., Borralleras, C., Pérez-Jurado, L.A. and Campuzano, V. TFII-I regulates target genes in the PI-3K and TGF- $\beta$  signaling pathways through a novel DNA binding motif. *Gene* 2013; 527 (2): 529-536.

Sheng, J., Luo, C., Jiang, Y., Hinds, P. W., Xu, Z., and Hu, G. F. Transcription of angiogenin and ribonuclease 4 is regulated by RNA Polymerase III elements and a CCCTC binding factor (CTCF)-dependent intragenic chromatin loop. *Journal of Biological Chemistry* 2014; 289 (18): 12520-12534.

Shipley, L.G. Assay design for Real-Time qPCR. In: Nolan, T. and Bustin, S.A., eds. 2013. *PCR Technology - Current Innovations*. Third ed. Florida, USA: CRC press (Taylor and Francis Group). 177 - 197.



Siegel, R., Naishadham, D., and Jemal, A. Cancer statistics CA: *A Cancer Journal for Clinicians* 2013; 63 (1): 11 - 30.

Singh, P., Lee, D. H., Szabó P., E. More than insulator: multiple roles of CTCF at the *H19 - Igf2* imprinted domain. *Frontiers in Genetics* 2012; 3: 214.

Skandalis, S.S., Afratis, N., Smirlaki, G., Nikitovic, D., Theocharis, A.D., Tzanakakis, G.N. and Karamanos, N.K. Cross-talk between estradiol receptor and EGFR/IGF-IR signaling pathways in estrogen-responsive breast cancers: Focus on the role and impact of proteoglycans. *Matrix Biology* 2013; 35:182-93.

Smith, I. E., Johnson, L., Dowsett, M., Robertson, J. F. R., Robison, L. E., Kokan, J. S., and Bliss, J. M. Trial of perioperative endocrine therapy: Individualizing care (POETIC). In *ASCO Annual Meeting Proceedings* 2011; 29 (15): TPS117).

Sobin LH, Gospodarowicz MK, Wittekind C (eds) 2010. *TNM Classification of Malignant Tumours* (7th Edition). American Joint Committee on Cancer (AJCC) / Union for International Cancer Control (UICC), Wiley-Blackwell, Oxford, UK.

Soto-Reyes E and Recillas-Targa F. Epigenetic regulation of the human *p53* gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene* 2010; 29: 2217 – 2227.

Soule, H. D., Vazquez, J., Long, A., Albert, S, and Brennan, M. A human cell line from a pleural effusion derived from a breast carcinoma. *Journal of the National Cancer Institute* 1973; 51(5): 1409-1416.

Splinter, E., Heath, H., Kooren, J., Palstra, R., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes and Development* 2006; 20 (17): 2349-2354.

Strom, A., Hartman, J., Foster, J.S., Kietz, S., Wimalasena, J. and Gustafsson, J.A. Estrogen receptor beta inhibits 17beta-estradiol-stimulated proliferation of the breast cancer cell line T47D. *Proceedings of the National Academy of Sciences of the United States of America* 2004; 101 (6): 1566 - 1571.

Stuart-Harris, R., Caldas, C., Pinder, S. E. and Pharoah, P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast* 2008; 17 (4): 323 - 334.

Swain, S., Kim, S., Cortes, J., Ro, J., Semiglazov, V., Campone, M., Ciruelos, E., Ferrero, J., Schneeweiss, A., Heeson, S., Clark, E., Ross, G., Benyunes, M. C., Baselga, J. Abstract 350O\_PR - Final overall survival (OS) analysis from the CLEOPATRA study of first-line (1L) pertuzumab (Ptz), trastuzumab (T), and docetaxel (D) in patients (pts) with HER2-positive metastatic breast cancer (MBC). European Society for Medical Oncology (ESMO) conference, Madrid, Spain. September 2014.

Szabó, P.E., Tang, S.H., Silva, F.J., Tsark, W.M., and Mann, J. R. Role of CTCF binding sites in the *Igf2* / *H19* imprinting control region. *Molecular and Cell Biology* 2004; 24: 4791 – 4800.

Tan, H., Zhong, Yili., and Pan, Z. Autocrine regulation of cell proliferation by estrogen receptor-alpha in estrogen receptor-alpha-positive breast cancer cell lines. *BMC Cancer* 2009; 9 (1): 31.

Tang, S. S., and Gui, G. P. Biomarkers in the diagnosis of primary and recurrent breast cancer. *Biomarkers in Medicine* 2012; 6 (5): 567-585.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X.S. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010; 329 (5991): 533-538.

Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., León, J., Klenova, E. and Delgado, M.D. CTCF regulates growth and erythroid differentiation of human myeloid leukaemia cells. *Journal of Biological Chemistry* 2005; 280 (30): 28152 – 28161.

Torrano, V., Navascués, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., León, J., Berciano, M.T., Renkawitz, R., Klenova, E., Lafarga, M. and Delgado, M.D. Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly (ADP-ribosyl)ation - dependent mechanism. *Journal of Cell Science* 2006; 119: 1746-1759.

Trinkle-Mulcahy, L. Resolving protein interactions and complexes by affinity purification followed by label - based quantitative mass spectrometry. *Proteomics* 2012; 12 (10): 1623-1638.

Tsai, C., Rowntree, R.K., Cohen, D.E. and Lee, J.T. Higher order chromatin structure at the X -inactivation centre via looping DNA. *Developmental Biology* 2008; 319 (2): 416-425.

Tsonis, A.I., Afratis, N., Gialeli, C., Ellina, M., Piperigkou, Z., Skandalis, S.S., Theocharis, A.D., Tzanakakis, G.N. and Karamanos, N.K. Evaluation of the coordinated actions of estrogen receptors with epidermal growth factor receptor and insulin-like growth factor receptor in the expression of cell surface heparan sulfate proteoglycans and cell motility in breast cancer cells. *FEBS Journal* 2013; 280 (10): 2248 - 2259.

Turner, P., McLellan, A., Bates, A. and White, M. eds., 2005. *Molecular Biology*. Third ed. USA and UK: Taylor & Francis Group.

Tutt, A., Robson, M., Garber, J. E., Domchek, S. M., Audeh, M. W., Weitzel, J. N. and Carmichael, J. Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and advanced breast cancer: a proof-of-concept trial. *Lancet* 2010; 376 (9737): 235-244.

Twelves, C., Cortes, J., Vahdat, L., Olivo, M., He, Y., Kaufman, P. A., and Awada, A. Efficacy of eribulin in women with metastatic breast cancer: a pooled analysis of two phases 3 studies. *Breast Cancer Research and Treatment* 2014; 148 (3): 553-61.

Urruticoechea, A., Smith, I.E. and Dowsett, M. Proliferation marker Ki-67 in early breast cancer. *Journal of Clinical Oncology* 2005; 23 (28): 7212 – 7220.

van Agthoven, T., Timmermans, M., Foekens, J.A., Dorssers, L. and Henzen-Logmans, S.C. Differential expression of estrogen, progesterone, and epidermal growth factor receptors in normal, benign, and malignant human breast tissues using dual staining immunohistochemistry. *American Journal of Pathology* 1994; 144 (6): 1238 - 46.

van Diest, P.J. Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology* 2004; 57 (7): 675.

Vermeulen, J. F., van de Ven, R. A., Ercan, C., van der Groep, P., van der Wall, E., Bult, P. and Derksen, P. W. (2012). Nuclear Kaiso expression is associated with high grade and triple-negative invasive breast cancer. *PloS one* 2012; 7 (5): e37864.

Viale, G., Giobbie-Hurder, A., Regan, M.M., Coates, A.S., Mastropasqua, M.G., Dell'Orto, P., Maiorano, E., MacGrogan, G., Braye, S.G. and Öhlschlegel, C. Prognostic and predictive value of centrally reviewed Ki-67 labelling index in postmenopausal women with endocrine-responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole. *Journal of Clinical Oncology* 2008; 26 (34): 5569-5575.

Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes and Development* 2009; 23 (22), 2563-2577.

Wallace, J.A. and Felsenfeld, G. We gather together: insulators and genome organization. *Current Opinion in Genetics and Development* 2007; 17(5): 400-407.

Wan, L., Pan, H., Hannenhalli, S., Cheng, Y., Ma, J., Fedoriw, A., Lobanenko, V., Latham, K.E., Schultz, R.M. and Bartolomei, M.S. Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Development* 2008; 135 (16): 2729-2738.

Wang, S., Nakajima, Y., Yu, Y., Xia, W., Chen, C., Yang, C., McIntush, E.W., Li, L., Hawke, D.H. and Kobayashi, R. Tyrosine phosphorylation controls PCNA function through protein stability. *Nature Cell Biology* 2006; 8 (12): 1359-1368.

Wang, Y., Yin, Q., Yu, Q., Zhang, J., Liu, Z., Wang, S. and Niu, Y. A retrospective study of breast cancer subtypes: the risk of relapse and the relations with treatments. *Breast Cancer Research and Treatment* 2011; 130 (2): 489-498.

Wang, D., Li, C., and Zhang, X. The promoter methylation status and mRNA expression levels of CTCF and SIRT6 in sporadic breast cancer. *DNA and Cell Biology* 2014; 33 (9): 581-90.

Waseem, N.H. and Lane, D.P. Monoclonal antibody analysis of the proliferating cell nuclear antigen (PCNA). Structural conservation and the detection of a nucleolar form. *Journal of Cell Science* 1990; 96 (1): 121-129.

Weaver, D. L., Ashikaga, T., Krag, D. N., Skelly, J. M., Anderson, S. J., Harlow, S. P., Juliam, B. T., Mamounas, E. P. and Wolmark, N. Effect of occult metastases on survival in node-negative breast cancer. *New England Journal of Medicine* 2011; 364 (5): 412-421.

Wells, A. EGF receptor. *International Journal of Biochemistry and Cell Biology* 1999; 31 (6): 637-643.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. and Peters, J. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 2008; 451 (7180): 796 - 801.

Witcher, M. and Emerson, B.M. Epigenetic silencing of the p16<sup>INK4a</sup> tumour suppressor is associated with loss of CTCF binding and a chromatin boundary. *Molecular Cell* 2009; 34 (3): 271-284.

Xia, X. Position weight matrix, gibbs sampler, and the associated significant tests in motif characterization and prediction. *Scientifica* 2012; 2012: 917540.

Xie, Z., Peng, J., Pennypacker, S.D. and Chen, Y. Critical role for the catalytic activity of phospholipase C- $\gamma$ 1 in epidermal growth factor-induced cell migration. *Biochemical and Biophysical Research Communications* 2010; 399 (3): 425 - 428.

Xu, N., Tsai, C. and Lee, J.T. Transient homologous chromosome pairing marks the onset of X inactivation. *Science* 2006; 311(5764): 1149-1152.

Xu, N., Donohoe, M.E., Silva, S.S. and Lee, J.T. Evidence that homologous X-chromosome pairing requires transcription and CTCF protein. *Nature Genetics* 2007; 39 (11): 1390-1396.

Xu J, Huo D, Chen Y, Nwachukwu C, Collins C, Rowell J, Slamon D, Olopade O. CpG island methylation affects accessibility of the proximal BRCA1 promoter to transcription factors. *Breast Cancer Research and Treatment* 2009; 120: 593 – 601.

Yang, X.J. and Chiang, C. M. Sumoylation in gene regulation, human disease and therapeutic action. *F1000Prime Reports* 2013; 5: 45.

Yager, J.D. and Davidson, N.E. Estrogen carcinogenesis in breast cancer. *New England Journal of Medicine* 2006; 354 (3): 270 - 282.

Yerushalmi, R., Woods, R., Ravdin, P.M., Hayes, M.M. and Gelmon, K.A. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncology* 2010; 11 (2): 174-183.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, A.P., Lobanenko, V., Klenova, E. and Ohlsson, R. Poly (ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nature Genetics* 2004; 36 (10): 1105-1110.

Zhang, R., Burke, L.J., Rasko, J.E., Lobanenko, V. and Renkawitz, R. Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Experimental Cell Research* 2004; 294 (1): 86-93.

Zhang, H., Niu, B., Hu, J. F., Ge, S., Wang, H., Li, T. and Hoffman, A. R. Interruption of intrachromosomal looping by CCCTC binding factor decoy proteins abrogates genomic imprinting of human insulin-like growth factor II. *Journal of Cell Biology* 2011; 193 (3): 475-487.

Zhao, H., Lo, Y., Ma, L., Waltz, S. E., Gray, J. K., Hung, M. C., and Wang, S. C. Targeting tyrosine phosphorylation of PCNA inhibits prostate cancer growth. *Molecular Cancer Therapeutics* 2011; 10 (1): 29 – 36.

Zhi, X., Wang, Y., Yu, J., Yu, J., Zhang, L., Yin, L. and Zhou, P. Potential prognostic biomarker CD73 regulates epidermal growth factor receptor expression in human breast cancer. *International Union of Biochemistry and Molecular Biology Life* 2012; 64 (11): 911-920.

Zhou, X., Werelius, B. and Lindblom, A. A screen for germline mutations in the gene encoding CCCTC - binding factor (CTCF) in familial non-BRCA1/BRCA2 breast cancer. *Breast Cancer Research* 2004; 6 (3): R187-R190.

Zhu, L., and Skoultschi, A. I. Coordinating cell proliferation and differentiation. *Current Opinion in Genetics and Development* 2001; 11 (1): 91 – 97.

Zlatanova, J. and Caiafa, P. CTCF and its protein partners: divide and rule? *Journal of Cell Science* 2009; 122: 1275-1284.

Zuckerman V, Wolyniec K, Sionov RV, Haupt S, Haupt Y. Tumour suppression by p53: the importance of apoptosis and cellular senescence. *Journal of Pathology* 2009; 219: 3 – 15.



## INTERNET SITES

Office for National Statistics, Cancer Statistics Registrations, England [Series MB1], accessed 18 December 2014, <<http://www.ons.gov.uk/ons/rel/vsob1/cancer-statistics-registrations--england--seriees-mb1-/index.html>>

Cancer Research UK, Stats Terminology and Calculations, accessed 14 December 2014, <<http://www.cancerresearchuk.org/cancer-info/cancerstats-explained/stats-terminology-and-calculations>>

WHO Fact sheet number 297. RDO-Cancer\_Fact\_Sheet\_n297\_feb2011.pdf., accessed 06 June 2013, <[www.afro.who.int](http://www.afro.who.int)>

WebPathology: Visual Survey of Surgical Pathology, accessed 15 January 2015, <http://www.webpathology.com/>

POETIC – Trial of Perioperative Endocrine Therapy – Individualising Care, accessed 8 Dec 2013, <http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=4023>

American Tissue Type Collection, accessed 4 January 2013, <http://www.lgcstandards-atcc.org>

Anti – Ki67 antibody (Sp6), accessed June 2012, <<http://www.abcam.com/ki67-sp6-antibody-ab16667.html>>

# APPENDIX

## 1. Reagents

Reagents	Source	Batch / lot number
RPMI with L Glutamine	Lonza	E84010-1918
Gentamicin	PAA	P00510-0560
Fetal Bovine Serum (FBS)	GIBCO	40F1043F
TRIS	Sigma	123K0181
Urea	Harnstoff	0918498
2 beta Mercaptoethanol	Sigma	34396CK
Protogel 30 (Acrylamide)	National Diagnostics	11 – 09 – 19
TEMED	Sigma	103K0656
Ammonium persulphate	Sigma	113K0656
Sodium dodecyl sulphate (SDS)	ACROS Organics	A0271956
Di-Sodium phosphate	Sigma	108K0174
Methanol	Fisher	1012778
Skimmed milk	Marvel	-
Tween-20	Sigma	038K0091

Citric acid	Sigma	058K0033
Dimethylsulfoxide (DMSO)	Sigma	034K0600
Bovine serum albumin (BSA) fraction V	Sigma	A9647-10G
Protein A Sepharose	Sigma	100M1421
HEPES	Fisher	074346
NaCl	Fisher	1153531
Na*EDTA	Fisher	1005265
PMSF	Sigma	054K2608
RNAasin	Fermentas	00082378
NP-40 (Igepal)	Sigma	Batch 034K0005
Sodium deoxycholate	Sigma	Batch 117K0124
EDTA	Acros Organics	A016735801
Pepstatin A	Sigma	010M8610V
PBS (phosphate buffered saline)	Fisher	096292
Glycin	Fisher	1086538
Coomasise blue	Fisher	Batch 9081
NaOH	Sigma	Batch 034K0109
Methanol	Fisher	1154574
Acetic acid	Fisher	Batch 1060086
Triton X	Acros Organics	A018186601

Sodium carbonate	Fisher	1141962
Formaldehyde	Sigma	F8775-25ML
Silver nitrate	Fisher	Batch 1007845
Sodium thiosulphate	Fisher	Batch 086602
Sodium acetate	Fisher	Batch 0890458
DMSO	Sigma	034K0600
HiMark prestained protein standard	Invitrogen	931966
Low molecular weight prestained protein ladder	NEB	P7703
4',6-diamidino-2-phenylindole (DAPI)	Thermo Scientific	Product # 622428
Developer (for Xray film)	Sigma	011M1091
Fixer (for Xray film)	Sigma	120M1129V
DharmaFECT 1	ThermoScientific	130820T
Attractene	Qiagen	Lot 142358039
RNA 6000 Nano dye	Agilent Technologies	1345
RNA 6000 Nano ladder	Agilent Technologies	1345
RNA 6000 gel matrix gel	Agilent Technologies	1345
96 – well plate	Applied Biosystems	4306737
Turbo DNase buffer (10x)	Ambion, USA	1202027
Turbo DNase (2U/μl)	Ambion, USA	1202019
Kapa mastermix	KapaBiosystems	KM4101

RNAse - free water	Fisher	127399
DNase inactivation reagent	Ambion, USA	Lot 1202053
Verso cDNA kit	ThermoScientific (USA)	00179194
Trisure	Bioline	BIO-38033
CTCF siRNA	ThermoScientific	140324
Non Target siRNA	ThermoScientific	1534236
Crosslink magnetic IP / co-IP kit	Pierce	NE 1742
Trypan blue (0.4%)	Life technologies	Cat 15250-061
Hanks Balanced Salt solution (HBSS)	Life technologies	Cat 14025092

## 2. Buffers / Gels / solutions

Buffers	Composition
Citrate buffer	10 mM citric acid pH 6.0 (adjusted with NaOH)
PBS / Glycin solution	PBS x1 (Fisher 096292) + 100 mM Glycin (Fisher 1086538)
2x lysis / loading buffer (1)	0.1M Tris / HCl pH 6.8; 7M Urea; 4% SDS; Phenol red dye (a pinch), 10% 2 $\beta$ -mercaptoethanol
2x lysis buffer (2)	1M TRIS / HCL pH 6.8, 10% SDS, glycerol, pinch of brilliant blue, 1M dithiothreitol (DTT)
Resolving buffer	1M Tris-HCl pH 6.8, 8.1% Acrylamide / Bis solution, 0.1% SDS; 10% ammonium persulphate (APS), TEMED 20ul for a 10ml gel solution.
Stacking buffer	0.1 M Tris-HCl pH 6.8, 4% Acrylamide / Bis solution, 0.1% SDS; 10% Ammonium Persulfate 50ul for a 10ml gel; TEMED 20ul for a 10ml gel solution
Running Buffer	3g TRIS, 15g Glycin, 10ml 10% SDS
Transfer Buffer	20 mM di-Sodium Phosphate pH 10-10.4; 0.05% SDS; 2% methanol
Blocking Solution	5% Tween; 1% BSA (optional); 50mM Tris pH 8.5 (buffered with citric acid)

Washing Buffer	PBS; Tween 0.05%
PBS / Milk	PBS, 3% dry milk powder
High stringency IP buffer  (IP lysis BUFFER 1)	BF1: 25 mM Tris/Hepes pH 8.0, 2 mM EDTA, 0.5% Tween,  1mM PMSF, 10mM pepstatin   BF2: 25 mM Tris/Hepes pH 8.0, 2 mM EDTA, 0.5% Tween,  1mM PMSF, 10mM pepstatin, 0.5 M NaCl
Medium stringency IP buffer  (IP lysis BUFFER 2)	LB-0.5: 50mM Tris/hepes pH 8.0, 0.5M NaCl, 2mM Na*EDTA, 1% NP-40, 1mM PMSF, 5 U/ml RNAsin.  LB-0: 50mM Tris/hepes pH 8.0, 2mM Na*EDTA, 1mM PMSF, 5 U/ml RNAsin  IB: 50mM Tris/hepes pH 8.0, 0.2M NaCl, 2mM Na*EDTA, 0.5% NP- 40, 1mM PMSF, 5 U/ml RNAsin
Low stringency IP buffer  (IP lysis BUFFER 3)	1M TRIS-HCL pH 7.4, 1% NP-40, 150 mM NaCl, 0.1M EGTA pH 8, 0.1M EDTA pH8, 1 mM PMSF, 2µg / µl Leupeptin, 2µg / µl aprotinin, Na Fluoride,  Na <sub>4</sub> P <sub>2</sub> O <sub>7</sub> – sodium pyrophosphate, Na <sub>2</sub> VO <sub>3</sub> – sodium orthovanadate
Sensitizer for silver staining	Sodium thiosulphate 2g; Sodium acetate 34g made up in 1L of solution with ice cold ultrapure water.
Silver nitrate 0.1%	1g silver nitrate in 1L of solution with ice cold ultrapure water
Developer for silver staining	Sodium carbonate 25g; Formaldehyde 37% - 200 µl; in 1L of solution with ice cold ultrapure water

Destain	40% methanol, 10% acetic acid, made up with distilled water
1% Agarose in Tris-Acetate-EDTA (TAE)	1g of agarose powder in 100ml TAE buffer
Luria broth (LB)	0.1% NaCl (Fisher, UK), 1% Bactotryptone (Fisher, UK) and 0.5% yeast extract (Fisher, USA) in distilled water.
Luria agar	2% bactoagar (Fisher, USA) in LB.
strip buffer composite solution	4g SDS, 1.4ml $\beta$ -mercaptoethanol, 1.51g TRIS in 200ml solution with distilled water
Solution 1	Tris-HCl, pH 8; 10mM EDTA; 100 $\mu$ g / ml RNase A
Solution 2	200 mM NaOH, 1% SDS
Solution 3	3M potassium acetate, pH 5.5



### 3. Antibodies

#### 3.1 Primary antibodies

Primary Antibody	Supplier / Product number	Application
Rabbit polyclonal anti-Nter (N terminal) CTCF (CTCF Nter)	In-house	IF, WB
Mouse monoclonal anti-human Ki67	Vector VP-K452 6001937	IF, WB, IP
Mouse monoclonal anti-PCNA	Abcam 29 960918	IF, WB, IP
Mouse monoclonal anti CTCF (CTCF BD)	BD Biosciences, 83171	IF, WB, IP
Rabbit polyclonal anti Ki 67	Abcam 833 GR29500-1	IF, WB, IP
Rabbit polyclonal anti-CTCF antibody	Millipore DAM 1772428	IP, WB, IP
Mouse monoclonal antibody to beta actin	Abcam 8224 GR14272-2	WB
Alpha RNA Pol II (N20) antibody	Santa Cruz , SC-899	WB

Mouse anti PARP 1 antibody	Enzo	ALX-210-302-R100	WB
----------------------------	------	------------------	----

### 3.2 Secondary antibodies

Secondary antibody	Supplier / product number	Application
Goat anti-rabbit secondary antibody conjugated to FITC	Abcam 6717	IF
Goat anti-mouse secondary antibody (Ig H+L) conjugated to TRITC	Southern Biotech K2007-Y957B	IF
Goat anti-mouse secondary conjugated to TRITC	Southern biotech 1080-03 A0011-Q141	IF
Goat anti-mouse secondary conjugated to FITC	Southern Biotech 1070-02 J5306 VM57	IF
Goat anti-rabbit (H &L) secondary antibody Horse Radish peroxidise (HRP) conjugated	Ab 6721	WB
Goat anti-mouse secondary HRP conjugated	Ab 6789 GR23382-2	WB

### 3.4 Biologic agents

Biologic agent	Supplier / product number	Application
<b>Plasmids</b>	In - house	Cell transfection
<b>E coli DH5<math>\alpha</math><sup>TM</sup> cells</b>	Life technologies	Bacterial transformation
<b>siRNA</b>  <b>(CTCF and Non target)</b>	ThermoScientific	Cell transfection

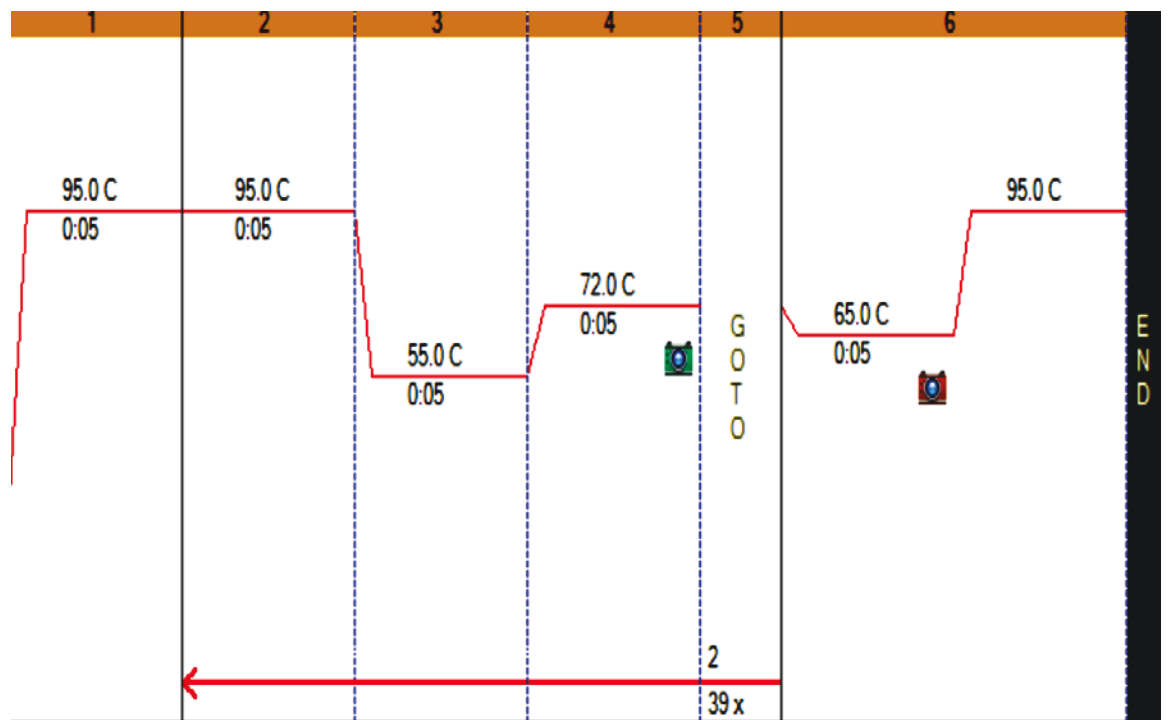
#### 4. QPCR primers

Batch #	Oligo Name	Oligo #	Len	Pur	Scale	MW	Tm°	µg/OD	OD	µg	nmol	Epsilon 1/(mMcm)	Dime2ndry	GC %	ul for 100µM	Sequence(5'-3')
HA04427511	ERS1 FWD	8013901487-000010	24	DST	0.025	7316	66.0	32.7	10.9	357.1	48.8	223.3	No	50	488	CTAAGTTCCTCTGGACAGGAACC
HA04427512	ERS1 REV	8013901487-000020	22	DST	0.025	6553	66.9	33.7	8.7	293.8	44.8	194	No	Very Weak	448	CCACATGCCCTCTACACATT
HA04427513	ERS2 FWD	8013901487-000030	21	DST	0.025	6464	66.2	31.7	8.4	267.0	41.3	203.3	No	Very Weak	413	AAGAAGATTCGCGGCTTTGTG
HA04427514	ERS2 REV	8013901487-000040	18	DST	0.025	5528	66.6	31.9	8.7	278.2	50.3	172.8	No	Very Weak	503	AAGCGCGCACTTGGTCG
HA04427515	PGR FWD	8013901487-000050	20	DST	0.025	6105	65.8	32.1	7.0	225.0	36.8	189.9	No	Very Weak	368	TCACTACCTGAGGCCGGAT
HA04427516	PGR REV	8013901487-000060	22	DST	0.025	6840	66.9	32.7	9.8	321.3	46.9	208.6	No	Weak	469	GAAGCTGATTTGGGCTCTGG
HA04427517	CTCF-1 FWD	8013901487-000070	24	DST	0.025	7318	64.9	29.9	10.4	311.1	42.5	244.6	No	None	425	AAACATTTACACGTGCGAATACCA
HA04427518	CTCF-1 REV	8013901487-000080	24	DST	0.025	7275	66.0	31.9	10.3	329.3	45.2	227.5	No	Very Weak	452	CCAGCACAATTATCAGCATGTCTT
HA04427519	CTCF-2 FWD	8013901487-000090	20	DST	0.025	6064	64.4	31.1	10.0	311.3	51.3	194.8	No	None	513	AGATCATGATTTCCAGCCCA
HA04427520	CTCF-2 REV	8013901487-000100	21	DST	0.025	6489	63.1	30.8	8.8	271.5	41.8	210.3	No	Very Weak	418	TGTGACAGTTTCATGTGCAAGA
HA04427521	BORIS-1 FWD	8013901487-000110	20	DST	0.025	6055	65.8	32.6	8.5	277.7	45.8	185.3	No	None	458	TGCAAAATTTTCATCCGACCTG
HA04427522	BORIS-1 REV	8013901487-000120	20	DST	0.025	6080	66.9	32.4	8.9	289.2	47.5	187.1	No	Weak	475	AAGCCTTTGCCACACTTGGAA
HA04427523	BORIS-2 FWD	8013901487-000130	20	DST	0.025	6120	66.0	32.2	9.1	293.1	47.8	190	Yes	Very Weak	50	CTGCTGCGGAACCATGTATA
HA04427524	BORIS-2 REV	8013901487-000140	21	DST	0.025	6415	67.5	32.4	9.4	305.0	47.5	197.7	No	Weak	476	TGCCATGTTGCAGTGGTACA

Batch #	Oligo Name	Oligo #	Len	Pur	Scale	MW	Tm°	µg/OD	OD	µg	nmol	Epsilon 1/(mMcm)	Dime2ndry	GC %	ul for 100µM	Sequence(5'-3')
HA04427452	CypB FWD	8013901719-000010	24	DST	0.025	7417	64.9	31.9	9.4	300.2	40.4	232.2	No	None	404	AGTGGATTAATTTTGTGGCCTTAGC
HA04427453	CypB REV	8013901719-000020	24	DST	0.025	7208	64.9	34.2	11.0	377.1	52.3	210.2	No	Very Weak	523	TTGTAGCCAAATCGTTCTCTCCT
HA04427454	TBP FWD	8013901719-000030	19	DST	0.025	5794	66.7	30.4	9.5	288.9	49.8	190.5	Yes	None	498	GCCCCAAGACGCCGAATATA
HA04427455	TBP REV	8013901719-000040	22	DST	0.025	6655	65.4	33.8	9.0	304.3	45.7	196.8	No	None	457	CGTGGCTCTTATCTCTCATGA
HA04427456	GUSB FWD	8013901719-000050	19	DST	0.025	5730	67.0	31.6	7.3	231.3	40.3	180.8	No	Weak	403	CCACACGGGACCATCCAAAT
HA04427457	GUSB REV	8013901719-000060	29	DST	0.025	8977	64.4	29.5	13.8	407.7	45.4	303.8	No	Weak	454	AGTCAAAATATGTCTTGGACAAAGTAA
HA04427458	GAPDH FWD	8013901719-000070	22	DST	0.025	6864	63.4	28.8	10.8	311.6	45.3	237.9	No	None	453	GCACAGAGGAGAGAGAGACC
HA04427459	GAPDH REV	8013901719-000080	20	DST	0.025	6296	64.0	30.6	10.9	334.4	53.1	205.2	No	None	531	AGGGGAGATTCAGTGTGGTG

## 5. QPCR temperature conditions (Kapa mastermix)



### 7.1 CTCF interacting partners identified in at least 2 breast cancer lines possessing 2 SC or more

CASP14	caspase 14, apoptosis-related cysteine peptidase;
GAPDH	glyceraldehyde-3-phosphate dehydrogenase;
C3	complement component 3; C3 plays a central role in the activation of the complement system
HIP1R	huntingtin interacting protein 1 related; Component of clathrin-coated pits and vesicles, that may link the endocytic machinery to the actin cytoskeleton. Binds 3-phosphoinositides (via ENTH domain).
ACLY	ATP citrate lyase;
DNAJC13	DnaJ (Hsp40) homolog, subfamily C, member 13
PPP1R12A	protein phosphatase 1, regulatory (inhibitor) subunit 12A
CTCF	CCCTC-binding factor (zinc finger protein);
MCM6	minichromosome maintenance complex component 6;
TF	transferrin;
GTF2H2	general transcription factor IIF, polypeptide 2, 44kDa; Component of the core-TFIIF basal transcription factor involved in nucleotide excision repair (NER) of DNA and, when complexed to CAK, in RNA transcription by RNA polymerase II. The N-terminus interacts with and regulates XPD whereas an intact C-terminus is required for a successful escape of RNAP II from the promoter (395 aa)

EIF4A1	eukaryotic translation initiation factor 4A, isoform 1;
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin);
UBTF	upstream binding transcription factor, RNA polymerase I;
TBC1D10C	TBC1 domain family, member 10C; Inhibits the Ras signaling pathway through its intrinsic Ras GTPase-activating protein (GAP) activity.
PABPC1	poly(A) binding protein, cytoplasmic 1;
SNX18	sorting nexin 18;
A2M	alpha-2-macroglobulin;
HSPA5	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa);
CAPRIN1	cell cycle associated protein 1
CTNNB1	catenin (cadherin-associated protein), beta 1,
FN1	fibronectin 1;
NUMA1	nuclear mitotic apparatus protein 1;
BSCL2	Berardinelli-Seip congenital lipodystrophy 2 (seipin) (462 aa)
BAG2	BCL2-associated athanogene 2
SERBP1	SERPINE1 mRNA binding protein 1
COLEC12	collectin sub-family member 12;)
Sep-09	septin 9; Filament-forming cytoskeletal GTPase (By similarity

## 7.2 KEY to CTCF - interacting proteins in *homo sapiens* generated by STRING

(Figure 3.19)

SUZ12	suppressor of zeste 12 homolog (Drosophila); Polycomb group (PcG) protein,
PARP1	poly (ADP-ribose) polymerase 1; Involved in the base excision repair (BER) pathway.
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin); Involved in diverse cellular processes
POLR2A	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa; DNA-dependent RNA polymerase]
SIN3A	SIN3 homolog A, transcription regulator (yeast); Acts as a transcriptional repressor.
YBX1	Y box binding protein 1; Binds to splice sites in pre-mRNA and regulates splice site selection.
POU5F1	POU class 5 homeobox 1 (360 aa)
Oct-04	POU domain, class 5, transcription factor 1 (Octamer-binding transcription factor 3)
PARG	poly (ADP-ribose) glycohydrolase
PARG99	Poly(ADP-ribose) glycohydrolase (EC 3.2.1.143)
RAD21	RAD21 homolog (S. pombe); Cleavable component of the cohesin complex.
SMAD3	SMAD family member 3; Transcriptional modulator activated by TGF-beta
GATA1	GATA binding protein 1 (globin transcription factor 1)
KCNQ5	potassium voltage-gated channel, KQT-like subfamily, member 5;



SYT8	synaptotagmin VIII; Not known.)
ZMYND10	zinc finger, MYND-type containing 10
SUMO2	SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae)
SMC3	structural maintenance of chromosomes 3; Involved in chromosome cohesion during cell cycle.
CHD8	chromodomain helicase DNA binding protein 8; DNA helicase that acts as a chromatin remodelling
SMC1A	structural maintenance of chromosomes 1A; Involved in chromosome cohesion during cell cycle
YY1	YY1 transcription factor; Multifunctional transcription factor
WSB1	WD repeat and SOCS box-containing 1; Probable substrate-recognition component of a SCF-like ECS [...] (421 aa)
KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1); Functions in nuclear protein import
HIST2H2AC	histone cluster 2, H2ac; Core component of nucleosome.

### 7.3 MCF7-specific protein partners with spectral count of 2 and over

DLD	dihydrolipoamide dehydrogenase;
HIP1R	huntingtin interacting protein 1 related
CDH1	cadherin 1, type 1, E-cadherin (epithelial);
LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein;
MAPK3	mitogen-activated protein kinase 3;
CTCF	CCCTC-binding factor (zinc finger protein
ATP6V1A	ATPase, H <sup>+</sup> transporting, lysosomal 70kDa, V1 subunit A;
GTF2H2	general transcription factor IIH, polypeptide 2, 44kDa;
SLC25A4	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator
HSPA4	heat shock 70kDa protein 4 (840 aa)
CDKN2AIP	CDKN2A interacting protein
PCBP1	poly(rC) binding protein 1
TBC1D10C	TBC1 domain family, member 10C
SNX18	sorting nexin 18
PKM2	pyruvate kinase, muscle;
CALR	calreticulin; Molecular calcium binding chaperone promoting folding
MDH2	malate dehydrogenase 2,

SRPR	signal recognition particle receptor (docking protein);
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5
ROBO	Rho GTPase activating protein 1;
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin
VCP	valosin-containing protein;
NCKAP1	NCK-associated protein 1
SMC3	structural maintenance of chromosomes 3
DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9
RFC3	replication factor C (activator 1) 3, 38kDa;
PFKP	phosphofructokinase, platelet (784 aa) 🌱
EWSR1	Ewing sarcoma breakpoint region 1
LDHA	lactate dehydrogenase A (332 aa)
MDM2	Mdm2 p53 binding protein homolog (mouse);

---

#### 7.4 T47D - specific protein partners with spectral count of 2 and over

MAZ MYC-associated zinc finger protein (purine-binding transcription factor);

C3	complement component 3;
HIP1R	huntingtin interacting protein 1 related;
APC	adenomatous polyposis coli; Tumor suppressor
CDH1	cadherin 1, type 1, E-cadherin (epithelial);
AXIN1	axin 1; Controls dorsoventral patterning via two opposing effects;
CTCF	CCCTC-binding factor (zinc finger protein
TF	transferrin
LEF1	lymphoid enhancer-binding factor 1;
CDH2	cadherin 2, type 1, N-cadherin (neuronal)
GTF2H2	general transcription factor IIH, polypeptide 2
ERCC3	Excision repair cross-complementing rodent repair deficiency,
NCSTN	nicastrin; Essential subunit of the gamma-secretase complex
UBTF	upstream binding transcription factor, RNA polymerase I;
TBC1D10C	TBC1 domain family, member 10C

JUP	junction plakoglobin; Common junctional plaque protein.
SNX18	sorting nexin 18; May be involved in several stages of intracellular trafficking
GSK3B	glycogen synthase kinase 3 beta; Participates in the Wnt signaling pathway.
HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1
PSEN1	presenilin 1; Probable catalytic subunit of the gamma-secretase complex,
CDH5	cadherin 5, type 2 (vascular endothelium)
CTNNB1	catenin (cadherin-associated protein), beta1
TFRC	transferrin receptor (p90, CD71);
LONP1	lon peptidase 1, mitochondrial; Required for intramitochondrial proteolysis
TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box
BTRC	beta-transducin repeat containing;
CTNNBIP1	catenin, beta interacting protein 1;
KIF20A	kinesin family member 20A;
CTNND1	catenin (cadherin-associated protein), delta 1;

---

#### 7.4 BT474 - specific protein partners with spectral count of 2 and over

CASP14	caspase 14, apoptosis-related cysteine peptidase;
C3	complement component 3
HIP1R	huntingtin interacting protein 1 related
LACRT	lacritin; Modulates secretion by lacrimal acinar cells (138 aa)
GTF2H4	general transcription factor IIH, polypeptide 4, 52kDa (462 aa)
CTCF	CCCTC-binding factor (zinc finger protein
GRIA2	glutamate receptor, ionotropic, AMPA 2; Ionotropic glutamate receptor
GTF2H2	general transcription factor IIH, polypeptide 2, 44kDa
GGCT	gamma-glutamyl cyclotransferase
GSDMA	gasdermin A; Induces apoptosis (445 aa)
TBC1D10C	TBC1 domain family, member 10C
SNX18	sorting nexin 18;
CR2	complement component (3d/Epstein Barr virus) receptor 2
CFH	complement factor H
CFI	complement factor I

## 7.5 MDA MB 231 - specific protein partners with spectral count of 2 and over

GAPDH	glyceraldehyde-3-phosphate dehydrogenase
C3	complement component 3;
HIP1R	huntingtin interacting protein 1 related;
GTF2H4	general transcription factor IIH, polypeptide 4, 52kDa (462 aa)
THBS1	thrombospondin 1
MED15	mediator complex subunit 15;
CTCF	CCCTC-binding factor (zinc finger protein);
GC	group-specific component (vitamin D binding protein);
CTDP1	CTD (carboxy-terminal domain
PAIP1	poly(A) binding protein interacting protein 1;
PABPC1	poly(A) binding protein, cytoplasmic 1;
GLS	glutaminase
G3BP1	GTPase activating protein (SH3 domain) binding protein 1
TGM2	transglutaminase 2
SERBP1	SERPINE1 mRNA binding protein 1
PGK1	phosphoglycerate kinase 1





## Table of Contents

<b>Introduction.....</b>	<b>1</b>
Figure 1.1. European age-standardised incidence rates per 100,000 population, by age, females, Great Britain.....	2
Figure 1.2. Breast Cancer mortality: 2010 – 2012, United Kingdom.....	5
1.1 Functional anatomy of the female breast.....	6
Figure 1.3 Anatomy of the normal female breast.....	7
Figure 1.4. Schematic representation of a transverse section through a normal breast duct.....	8
1.2 Classification of Breast Cancer.....	9
Figure 1.5. Histology of breast ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS). ....	11
Figure 1.6. Histology of invasive ductal (IDC) and invasive lobular carcinoma (ILC). ....	12
Figure 1. 7. Human breast tumour subtypes and possible link to normal breast epithelial hierarchy. ....	13
1.3 Prognostic factors in breast cancer.....	16
1.5 Epigenetic alterations in breast cancer.....	21
1.6 The CCCTC-binding factor (CTCF).....	23
1.6.1 History, discovery and conservation of CTCF.....	23
1.6.2 CTCF structure.....	24
Figure 1.8. Structural features of CTCF.....	26
Figure 1.9. Schematic structure of the classic Cys2 His2 zinc finger (ZF)....	27
1.6.3 CTCF functions.....	28
Figure 1.10. CTCF and imprinting at the Igf2 / H19 gene locus via loop formation. ....	33
1.7 CTCF DNA – binding sites.....	35
1.7.1 Distribution and characteristics of CTCF DNA - binding sites.....	35
1.7.2 Classification and functional effect of CTCF binding sites.....	37
1.8 CTCF protein partners.....	38
Figure 1.11. Schematic representation of classes of CTCF protein partners. ....	39
1.8.1 Chromatin CTCF - protein partners.....	40
1.8.2 DNA-binding CTCF protein partners.....	40
1.8.3 Multifunctional CTCF protein partners.....	41
1.8.4 Miscellaneous CTCF protein partners.....	42
1.9 Regulation of CTCF activity.....	42
1.9.1 Methylation.....	43
1.9.2 Post translational modifications (PTM).....	43

1.10 CTCF and tumour suppression.....	45
1.10.1 CTCF induces transcriptional repression of hTERT in tumours .....	45
1.10.2 CTCF maintains retinoblastoma protein (pRb) and p53 gene promoter .....	46
epigenetic status and tumour repression.....	46
1.10.3 CTCF maintains epigenetic balance at the cyclin-dependent kinase inhibitor .....	47
2A locus (CDKN2A) .....	47
1.11 CTCF and the clinical breast cancer phenotype.....	48
1.12 CTCF and the estrogen receptor (ER) $\alpha$ .....	51
2.1 MATERIALS .....	55
2.1.1 Breast cancer cell (BCC) cell lines .....	55
2.1.2 Culture media .....	55
2.1.2.1 Culture medium for MCF7, T47D and BT474 cells.....	55
2.1.3 Reagents and Buffers /gels / solutions .....	56
2.1.4 Antibodies .....	56
2.1.5 Plasmids, siRNA and biologic agents .....	56
2.1.6 QPCR primers.....	56
.....	57
2.2 METHODS.....	58
2.2.1. Cell culture procedures.....	58
2.2.2 Trypan blue test for cell viability .....	59
2.2.3 Breast cancer cell lysates .....	59
2.2.4 Bovine serum albumin protein assay .....	60
2.2.5 Indirect Immunofluorescence procedure .....	60
2.2.6 Sodium Dodecyl Sulphate (SDS) – PolyAcrylamide Gel Electrophoresis (PAGE) and western blot analysis .....	62
2.2.7 Immunoprecipitation assay (IP).....	65
2.2.8 RNA – based procedures .....	66
2.2.9 Plasmid DNA procedures.....	68
2.2.10 Transfection assays .....	71
2.2.11 Reverse transcription – polymerase chain reaction for (RT-PCR; QPCR) procedures.....	73
2.2.12 Liquid chromatography – mass spectrometry (LC – MS / MS).....	74
<b>RESULTS - Investigating CTCF protein partners in a panel of five breast cancer cell lines .....</b>	<b>75</b>
3.1 Background .....	75
3.1.1 Cell proliferation and breast cancer.....	75

3.1.2 CTCF and proliferation .....	75
3.1.3 Ki67 protein and proliferation.....	77
3.1.4 Proliferating cell nuclear antigen (PCNA), proliferation and cancer .....	79
3.2 Knowledge gap and hypothesis.....	80
3.3 Objectives of this chapter.....	80
3.4 Results: Novel CTCF protein partners in five breast cancer cell lines .....	81
3.4.1 Confirmation of estrogen receptor (ER), progesterone receptor (PR) and HER2 receptor expression status in breast cancer cell lines .....	81
Figure 3.1: Hormone receptor (ER and PR) and HER2 expression profile of five breast cancer cell lines.....	82
3.4.2 Total CTCF protein expression in different breast cancer cell lines.....	83
Figure 3.2: Total CTCF protein expression in a panel of five breast cancer cell lines and one normal breast epithelial cell line.....	85
3.4.3 Differential expression of proliferation markers, Ki67 and PCNA, in a panel of different breast cancer cell lines .....	86
Figure 3.3: Expression of proliferation markers Ki67 and PCNA in a panel of five different breast cancer cell lines and one normal breast epithelial cell line.....	88
3.4.4 CTCF localisation in relation to breast cancer phenotype and anti- CTCF antibody type .....	89
Figure 3.4. Single indirect immunofluorescence staining of five breast cancer cell lines with CTCF primary antibodies.....	91
3.4.5 Ki67 protein localisation in a panel of five breast cancer cell lines using two different anti Ki67 antibodies .....	93
.....	94
Figure 3.5. Single immunofluorescence staining of five breast cancer cell lines with Ki67 primary antibodies. ....	94
3.4.6 PCNA protein localisation localisation in breast cancer cell lines.....	97
Figure 3.6. Single immunofluorescence staining of five breast cancer cell lines with PCNA primary antibody. ....	98
3.4.7 Assessment of immunofluorescence bleed through .....	100
Figure 3.7. Assessment of fluorophore-conjugated secondary antibody bleedthrough .....	101
3.4.8 CTCF protein co-localisation with Ki67 protein in breast cancer cell lines.....	102
Figure 3.8. A – E. Double immunofluorescence of CTCF protein with Ki67 protein in a panel of five breast cancer cell lines.....	103
3.4.9 CTCF protein colocalisation with PCNA protein in breast cancer cell lines with different hormone receptor / HER2 and invasive properties.....	106
Figure 3.9 A - E. Double indirect immunofluorescence of CTCF with PCNA in a panel of five breast cancer cell lines.....	107

3.4.10 CTCF immunoprecipitation (IP) and coimmunoprecipitation (co-IP) with Ki67 and PCNA in MCF7 breast cancer cells using a high stringency IP buffer.....	111
Figure 3.10. Silver stain (A), CTCF immunoprecipitation (B) and CTCF co-immunoprecipitation (C) in MCF7 cells using a high stringency buffer.....	113
3.4.11 CTCF co-immunoprecipitation with Ki67, PCNA and known protein partners (RNA pol II and PARP 1) in MCF7 breast cancer cells using a medium stringency buffer.....	114
Figure 3.11. CTCF co-IP with Ki67, PCNA, RNA pol II, and PARP 1 in MCF7 breast cancer cells using a medium stringency IP buffer.....	115
3.4.12. CTCF immunoprecipitation and coprecipitation with Ki67, PCNA and known protein partners (RNA pol II and PARP 1) in MCF7 breast cancer cells using a low stringency buffer.....	116
Figure 3.12. CTCF IP and co-IP with Ki67 / PCNA and known protein partners RNA pol II / PARP 1, in MCF7 breast cancer cells using a low stringency buffer .....	118
3.4.13. PCNA immunoprecipitation and coprecipitation with CTCF and Ki67 in the MCF7 breast cancer cell line using a low stringency IP lysis buffer..	119
Figure 3.13. PCNA immunoprecipitation and coprecipitation with CTCF / Ki67 in MCF7 breast cancer cells using a low stringency IP lysis buffer.....	120
3.4.14 Immunoprecipitation of CTCF and coprecipitation with Ki67 / PCNA using a cross-linker and a magnetic commercial kit .....	121
Figure 3.14. CTCF immunoprecipitation (IP) and co-immunoprecipitation (co-IP) in MCF7 cells using a magnetic co-IP kit incorporating a crosslinker. ....	122
3.4.15 RNA pol II immunoprecipitation and coprecipitation with CTCF in the HeLa cervical cancer cell line cell using a low stringency IP lysis buffer (buffer 3).....	123
Figure 3.15. RNA pol II and CTCF coimmunoprecipitation in HeLa cell extracts.....	124
3.4.16 Breast cancer cell line - specific differences in CTCF coprecipitation with Ki67 and PCNA in breast cancer cell lines.....	125
Figure 3.16. A - D. CTCF IP and co-IP with Ki67 / PCNA in T47D, BT474, SKBR3 and MDA MB 231 breast cancer cell lines .....	126
3.4.17 CTCF protein interaction with Ki67 / PCNA assessed by Liquid Chromatography - Mass Spectrometry (LC – MS / MS) in breast cancer cell lines.....	128
Figure 3.17. CTCF spectral scores on liquid chromatography / mass spectrometry in breast cancer cell lines.....	130
3.4.18 Novel CTCF-interacting partners in breast cancer cell lines .....	132
Figure 3.18. CTCF-interacting proteins identified across all cell lines .....	134
3.4.19 CTCF interacting protein network map in breast cancer cell lines ...	135

Figure 3.19. Action view of CTCF interacting proteins identified in at least two breast cancer cell lines with SC of 2 or more .....	137
Figure 3.20. STRING output of known and predicted CTCF interacting protein partners in homo sapiens. ....	138
Figure 3.21 A – D. Action view of breast cancer cell line - specific CTCF protein interaction network .....	139
<b>Discussion: Novel CTCF interacting partners</b> .....	143
<b>Regulatory relationship between CTCF and ER<math>\alpha</math></b> .....	158
<b>5 Background</b> .....	158
5.1 Role of estrogen in breast cancer initiation .....	158
Figure 5.1. Inter-relation of estrogen, ER and EGFR receptor signalling cascades .....	159
5.2 ER structure and activation .....	161
Figure 5.2. Structure and comparison of estrogen $\alpha$ and $\beta$ receptor domains .....	162
5.3 ER $\alpha$ regulation .....	163
5.4 Knowledge gap and hypothesis .....	165
5.5 Objectives of this chapter .....	165
5.6 Results .....	166
5.6.1 Restriction enzyme digest of CTCF plasmid expression vectors .....	166
Figure 5.3. Restriction enzyme digest of CTCF pCI expression vector .....	168
Figure 5.4. Restriction enzyme digest of the pCI empty vector (EV). ....	169
5.6.2 Optimisation of CTCF plasmid DNA overexpression assays in MCF7 cells .....	170
Figure 5.5. Optimisation of MCF7 overexpression with a plasmid expression vector .....	172
5.6.3 Assessment of siRNA transfection efficiency and experimental set up with the positive control cyclophilin siRNA .....	173
Figure 5.6. CTCF siRNA transfection efficiency in MCF7 cells with cyclophilin B positive control. ....	174
5.6.4 Optimisation of CTCF siRNA knockdown assays in MCF7 cells .....	175
Figure 5.7. MCF7 cell RNA interference (RNAi) with CTCF and non target siRNA .....	177
5.6.5 Effect of transfection agents and reagents on MCF7 cell growth and viability .....	178
Figure 5.8. MCF7 cell response to expression vectors and attractene transfection reagent. ....	180
Figure 5.9. MCF7 cell response to siRNA and DharmaFECT transfection reagent .....	181
5.6.6 QPCR measurements of CTCF and ER $\alpha$ gene expression following plasmid overexpression and siRNA knockdown assays in MCF7 cells .....	182

Figure 5.10. Densitometry plot of RNA quality assessment using the Agilent Bioanalyser.....	185
Figure 5.11. Spherograms of RNA quality assessment using the Agilent Bioanalyser.....	186
Figure 5.12. Nucleic acid primer pair fold pattern for CTCF and ER $\alpha$ primer pairs. ....	189
Figure 5.13. Nucleic acid primer pair fold pattern for TBP and GAPDH primer pairs ....	190
Figure 5.14. Testing primers: amplification plots for linear range .....	192
Figure 5.15. Standard curves for QPR efficiency .....	196
Figure 5.16. Assessment of QPCR replicate consistency and melting curves.....	198
Figure 5.17. CTCF and ER $\alpha$ mRNA fold change on CTCF overexpression.....	201
Figure 5.18. CTCF and ER mRNA fold change on siRNA knockdown. ....	203
Discussion – Regulatory effect of CTCF on ER $\alpha$ expression.....	204
Final discussion and conclusions .....	209
FUTUTRE DIRECTIONS .....	213
REFERENCES.....	214
Beneke S. Regulation of chromatin structure by poly(ADP-ribosyl)ation. <i>Frontiers in Genetics</i> 2012; 3: 169. ....	215
Bilal, E., Dutkowski, J., Guinney, J., Jang, I. S., Logsdon, B. A., Pandey, G. and Margolin, A. A. Improving breast cancer survival analysis through competition-based multidimensional modeling. <i>PLoS Computational Biology</i> 2013; 9 (5): e1003047.....	215
Celis, J.E. and Celis, A. Cell cycle-dependent variations in the distribution of the nuclear protein cyclin proliferating cell nuclear antigen in cultured cells: subdivision of S phase. <i>Proceedings of the National Academy of Sciences</i> 1985; 82 (10): 3262 - 3266. ....	217
Dundr, M., Misteli, T. and Olson, M.O. The dynamics of post mitotic reassembly of the nucleolus. <i>Journal of Cell Biology</i> 2000; 150 (3): 433-446.....	220
Fox, J.T., Lee, K. and Myung, K. Dynamic regulation of PCNA ubiquitylation / deubiquitylation. <i>FEBS Letters</i> 2011; 585 (18): 2780-2785. ....	222
Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A. and Jensen, L. J. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. <i>Nucleic Acids Research</i> 2013; 41 (D1): D808 - D815.....	222
INTERNET SITES.....	249
APPENDIX .....	250
<b>1. Reagents</b> .....	250
Cat 15250-061 .....	253
2. Buffers / Gels / solutions.....	254
3. Antibodies .....	257

3.1 Primary antibodies.....	257
3.2 Secondary antibodies.....	258
3.4 Biologic agents.....	259
4. QPCR primers .....	260
5. QPCR temperature conditions (Kapa mastermix) .....	261

